(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number
WO 02/070699 A2

(51) International Patent Classification[7]: C12N 15/12, 5/10, C07K 14/47, 16/18, G01N 33/68, A01K 67/027

(21) International Application Number: PCT/EP02/02189

(22) International Filing Date: 28 February 2002 (28.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01105062.2 1 March 2001 (01.03.2001) EP

(71) Applicant *(for all designated States except US)*: **LION BIOSCIENCE AG** [DE/DE]; Im Neuenheimer Feld 515, 69120 Heidelberg (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ALBERS, Michael** [DE/DE]; Obere Schulstrasse 3, 69221 Dossenheim (DE). **ELLWANGER, Silvia** [DE/DE]; Theodor-Heuss-Strasse 4, 69181 Leimen (DE). **LÖSER, Eva** [DE/DE]; Bieth-strasse 12, 69121 Heidelberg (DE). **KÖGL, Manfred** [AT/DE]; Hauptstrasse 131/4, 69214 Eppelheim (DE).

(74) Agents: **GODDAR, Heinz** et al.; Boehmert & Boehmert, Pettenkoferstrasse 20-22, 80336 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/070699 A2

(54) Title: NOVEL COFACTORS OF THE ESTROGEN RECEPTOR ALPHA AND METHODS OF USE

(57) Abstract: The present invention relates to novel cofactors of the estrogen receptor alpha which are designated CF16, CF17, CF18, CF19, CF40, CF41, CF42 and CF43 and in particular to the isolated nucleic acid sequences encoding these cofactors and the isolated polypeptides thereof. The invention further relates to processes for isolating and/or producing the nucleic acids or the proteins as well as methods of use of these cofactors, such as inhibiting or activating the binding of the cofactors to the estrogen receptor alpha.

NOVEL COFACTORS OF THE ESTROGEN RECEPTOR ALPHA AND METHODS OF USE

BACKGROUND OF THE INVENTION

Multicellular organisms are dependent on advanced mechanisms of information transfer between cells and body compartments. The information that is transmitted can be highly complex and can result in the alteration of genetic programs involved in cellular proliferation, differentiation or reproduction. The signals, such as hormones are often simple molecules, such as peptides, fatty acids, or cholesterol derivatives.

Many of these signals produce their effects by ultimately changing the transcription of specific genes. One well-studied group of polypeptides that mediate a cell's response to a variety of signals is a family of transcription factors known as nuclear receptors, hereinafter referred to frequently as "NR". Members of this group include receptors for steroid hormones (for example, estrogens and glucocorticoids and other cholesterol-derivatives), vitamin D, ecdysone, cis and trans retinoic acid, thyroid hormone, bile acids, fatty acids (and other peroxisomal proliferators), as well as so-called orphan receptors, proteins that are structurally similar to other members of this group, but for which no ligands are known (Escriva, H. et al., Ligand binding was acquired during evolution of nuclear receptors, PNAS, 94, 6803 – 6808, 1997). Orphan receptors may be indicative of unknown signaling pathways in the cell or may be nuclear receptors that function without ligand activation. The activation of transcription by some of these orphan receptors may occur in the absence of an exogenous ligand and/or through signal transduction pathways originating from the cell surface (Mangelsdorf, D. J. et al., The nuclear receptor superfamily: the second decade, Cell 83, 835-839, 1995).

In general, three functional domains have been defined in NRs. An amino terminal domain is believed to have some regulatory function. A DNA-binding domain hereinafter referred to as "DBD" usually comprises two zinc finger elements and recognizes a specific Hormone Responsive Element hereinafter referred to as "HRE" within the promoters of responsive genes. Specific amino acid residues in the "DBD" have been shown to confer DNA sequence binding specificity (Schena, M. & Yamamoto, K.R., Mammalian Glucocorticoid Receptor Derivatives

Enhance Transcription in Yeast, Science, 241:965-967, 1988). A Ligand-binding-domain hereinafter referred to as "LBD" is at the carboxy-terminal region of known NRs. In the absence of hormone, the LBD appears to interfere with the interaction of the DBD with its HRE. Hormone binding seems to result in a conformational change in the NR and thus opens this interference (Brzozowski et al., Molecular basis of agonism and antagonism in the oestogen receptor, Nature, 389, 753 – 758, 1997; Wagner et al., A structural role for hormone in the thyroid hormone receptor, Nature, 378, 690 – 697. 1995). A NR without the LBD constitutively activates transcription but at a low level.

Both the amino-terminal domain and the LBD of the NR appear to have transcription activation functions hereinafter referred to as "TAF". Acidic residues in the amino-terminal domains of some nuclear receptors may be important for these transcription factors to interact with RNA polymerase. TAF activity may be dependent on interactions with other protein factors or nuclear components (Diamond et al., Transcription Factor Interactions: Selectors of Positive or Negative Regulation from a Single DNA Element, Science, 249:1266-1272 , 1990). Certain oncoproteins (e.g., c-Jun and c-Fos) can show synergistic or antagonistic activity with glucocorticoid receptors (GR) in transfected cells. Furthermore, the receptors for estrogen, vitamins A and D, and fatty acids have been shown to interact, either physically or functionally, with the Jun and Fos components of AP-1 in the transactivation of steroid- or AP-1 regulated genes.

Coactivators of transcription are proposed to bridge between sequence specific transcription factors, the basal transcription machinery and in addition to influence the chromatin structure of a target cell. Several proteins like SRC-1, ACTR, and Grip1, which are also cofactors of NRs similar to those disclosed in this invention, interact with NRs in a ligand enhanced manner (Heery et al., A signature motif in transcriptional coactivators mediates binding to nuclear receptors, Nature, 387, 733 – 736; Heinzel et al., A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression, Nature 387, 43 – 47, 1997). Furthermore, the physical interaction with negative receptor-interacting proteins or corepressors has been demonstrated (Xu et al., Coactivator and Corepressor complexes in nuclear receptor function, Curr Opin Genet Dev, 9 (2), 140 – 147, 1999).

Nuclear receptor ligands like steroid hormones affect the growth and function of specific cells by binding to intracellular receptors and forming nuclear receptor-ligand complexes. Nuclear

receptor-hormone complexes then interact with a hormone response element (HRE) in the control region of specific genes and alter specific gene expression.

The present invention relates to the identification of novel interacting proteins of the estrogen receptor alpha.

The estrogen receptors are nuclear steroid receptors that mediate the effects of estrogen in the body and are therefore involved in the regulation of important developmental and physiological processes such as sexual differentiation and behaviour, fertility, cardiovascular function, brain function, bone generation and resorption as well as cell proliferation and carcinogenesis.

Estrogen receptors exist in two isoforms, which are encoded on two separate genes. The two isoforms, termed estrogen receptor alpha and estrogen receptor beta (hereinafter referred to as ER alpha and ER beta, respectively) share some degree of structural and functional similarity. However, differences with respect to structure and tissue expression patterns have been recognised which suggest that the two estrogen receptors fulfil distinct physiological roles in many tissues.

It has been shown that ligands exist or have been chemically designed (both agonists and antagonists), which selectively modulate the action of only one of the two isoforms, thereby opening ways to more specifically treat medical indications influenced by the action of estrogen (reviewed in Katzenellenbogen et al., Recent Prog Horm Res 55, 163-193 (2000) and Barkhem et al., Mol Pharmacol 54, 105-12 (1998)).

Although both the alpha and beta isoforms are expressed in a range of tissues such as the central nervous system, the cardiovascular system, the immune system, the urogenital tract, the gastrointestinal tract, the bone, the lungs, the mammary gland and the uterus, expression of one isoform can be predominant in some cell types. For instance, expression of ER alpha in the adult uterus and in the mammary glands is more pronounced than ER beta expression, whereas in the urogenital tract, ER beta seems to be the physiological important form (reviewed in Gustafsson, J Endocr 163, 379-383 (1999)).

ER alpha seems to be responsible for most of estrogen's effects on reproduction and reproductive organs, which are fully compromised in its absence in adult female mice (Lubahn et

al., PNAS 90, 11162-11166 (1993)). Females are infertile with hypoplastic uteri and hyperemic ovaries and they lack breast tissue development. Males are also infertile. However, despite the fact that ER beta expression in the uterine tissue is low and despite the fact that in ER alpha knock out mice the uterotrophic response to estrogen is diminished, also mice disrupted for the ER beta locus (BERKOs) show a decreased reproductive performance, suggesting a requirement for ER beta for full reproductive capability (Krege et al., PNAS 95, 15677-15682 (1998)).

It has long been proposed that elevated estrogen levels might increase in the breast cancer risk in postmenopausal women. Evidence has been put forward that this is at least in part due to an influence of estrogens on the activity of the breast cancer susceptibility gene BRCA1. In turn it has been shown that the activity of the estrogen receptor alpha can be suppressed in transfected cells by BRCA1 (Fan et al., Science 284, 1354-56 (1999)). Mutations in the BRCA1 gene or its impaired function in older women (e.g. through increased methylation associated with ageing) might thus lead to a decreased suppression of the proliferative functions of the estrogen receptor on mammary epithelial cells and as a consequence to an increased level developing breast tumors. Therefore, substances or cofactors that inhibit the activity of the estrogen receptor alpha in mammary cells are potential candidates for the prevention of breast cancer.

In non-reproductive tissues the estrogen receptors are implicated in the maintenance of bone mineral density and cardiovascular health in women. Administration of estrogen and a class of drugs referred to as selective estrogen receptor modulators (SERMs) have since long been considered as the first line therapy for osteoporosis in postmenopausal women. Estrogens inhibit osteoclast generation thereby reducing the resorption of bone material (reviewed in Rodan et al., Science 289, 1508-1524 (2000)). SERMs, such as tamoxifene and raloxifene, bind with high affinity to estrogen receptors. It seems that each SERM bound to ER forces the receptor into a distinct conformation, allowing the recruitment of a specific set of cofactor proteins (which will either be activating or repressing) in a tissue dependent manner. For instance, raloxifene operates as an agonist in bone but as an antagonist in breast and uterus and thus can be applied to prevent osteoporosis and furthermore to reduce the risk of breast cancer in postmenopausal women by opposing the effects of circulating estrogen.

Some data suggest a specific role for ER beta in mediating cardiovascular effects of estrogen, since estrogen protects against vascular injury in mice deficient in ER alpha (Iafrati et al., Nat Med 3, 545 (1997)). Furthermore, expression of ER beta, but not of ER alpha, is markedly increased in vascular cells after vascular injury (Lindner et al., Circ Res 83, 224 (1998); Makela et al., PNAS 96, 7077 (1999). Thus, the protective effect of estrogen on vascular lesions might be mediated by ER beta involving inhibition of smooth muscle cell proliferation.

Recent data propose mechanisms by which ER beta might exert its regulatory function on cell proliferation and its protective function against cancer. Montano et al., J Biol Chem 273, 25443-25449 (1998) show that antiestrogens are able to induce expression of the quinone reductase (QR) gene in breast cancer cells and that binding of antiestrogen-liganded ER beta to an antioxidant response element in the promoter of the gene is required for this induction. Transcriptional activation of the QR gene by antiestrogen-liganded ER alpha was much less pronounced.

Thus, antioxidant-regulated genes, such as the QR gene, which products control the concentrations of free radicals and reactive oxygen - important players in the onset and course of cancer - might be regulated by ER beta. An alternative way, in which cell proliferation could be controlled by ER beta was suggested by Poelzl et al., PNAS 97, 2836-2839 (2000). Here, ER beta, but not ER alpha, was demonstrated to interact directly and specifically with a cell-cycle regulatory protein, MAD2 (mitosis arrest-deficient 2) in a ligand independent manner. This could suggest, that the regulatory functions of ER beta in cell proliferation might be mediated through direct protein-protein contacts with a cell cycle spindle assembly protein and thus in a way distinct from the established function of the ERs as transcription factors.

Although as described above both ERs seem to differ in their mode of action it should be pointed out that it is known that, for instance in cells of the hypothalamus (Pettersson et al., Mol Endocrin 11, 1486-96 (1997)), ER alpha and ER beta form heterodimers and thus might be able to even regulate each other directly.

The present invention relates to the identification of novel interacting polypeptides of the estrogen receptor alpha. The identification and characterisation of protein factors which modulate ER transactivation activity could be of great benefit for the treatment of numerous diseases such as osteoporosis and other bone diseases, failures in reproductive functions, cancer,

cardiovascular diseases such as atherosclerosis, as well as the prevention of hot flushes, mood changes and Alzheimer's disease.

The present invention provides novel proteins, nucleic acids, and methods useful for developing and identifying compounds for the treatment of these diseases. The invention also provides for methods to test if a certain compound promotes or disrupts the interaction of these proteins with ER alpha, allowing the screening for compounds with estrogen-regulated cellular effects. These novel proteins interact, presumably also *in vivo*, with the ER alpha receptor and shall hereinafter collectively be referred to as "cofactors" or "CFs", although some of them in fact do belong to the nuclear receptor family of polypeptides.

The importance of this invention is manifested in the effects of the CFs to modulate genes involved in cellular functions like regulation of metabolism and cell homeostasis, cell proliferation and differentiation, pathological cellular aberrations, or cellular defense mechanisms.

The CF proteins are useful for screening for ligands of the ER alpha thereby providing for agents which influence the activity of ER alpha and thus the activity of genes controlled by ER alpha.

In one aspect of the present invention, the present invention provides isolated nucleic acid sequences for novel CFs. In particular, the present invention provides the cDNA sequences encoding human CFs.

These nucleic acid sequences have a variety of uses. For example, they are useful for making vectors and for transforming cells, both of which are ultimately useful for production of the CF polypeptides. They are also useful as scientific research tools for developing nucleic acid probes for determining expression levels of the cofactor genes, e.g., to identify diseased or otherwise abnormal states. They are useful for developing analytical tools such as anti sense oligonucleotides for selectively inhibiting expression of the cofactor genes to determine physiological responses.

In another aspect of the present invention, a homogenous composition comprising the cofactor proteins is provided. The protein is useful for screening drugs for agonist and antagonist activity, and, therefore, for screening for drugs useful in regulating physiological responses

associated with the cofactors according to the invention. Specifically, antagonists to the CFs could be used to treat metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases. The proteins are also useful for developing antibodies for detection of the proteins.

Flowing from the foregoing are a number of other aspects of the invention, including (a) vectors, such as plasmids, comprising the cofactor nucleic acid sequences that may further comprise additional regulatory elements, e.g., promoters, (b) transformed cells that express the cofactors, (c) nucleic acid probes, (d) antisense oligonucleotides, (e) agonists, (f) antagonists, and (g) transgenic mammals. Further aspects of the invention comprise methods for making and using the foregoing compounds and compositions.

The problem underlying the present invention is thus solved by the independent claims of the attached set of claims. Embodiments thereof can be taken from the subclaims.

The foregoing merely summarizes certain aspects of the present invention and is not intended, nor should it be construed, to limit the invention in any manner. All patents and other publications recited herein are hereby incorporated by reference in their entirety.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

THE CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43 POLYPEPTIDES AND THEIR RESPECTIVE NUCLEIC ACIDS:

The present invention comprises, in part, novel cofactors (CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43) of the mammalian ER alpha. Particularly preferred embodiments of these cofactors are those having an amino acid sequence substantially the same as SEQ ID NOs. 3, 6, 9, 12, 15, 18, 21, and/or 24.

As used herein, if reference to the "cofactor" is made, it is meant as a reference to any protein having an amino acid sequence substantially identical to any of SEQ ID NOs. 3, 6, 9, 12, 15, 18, 21, and/or 24.

As used herein, if reference to the cofactor is made or the cofactor "X",wherein "X" stands for the number designating the cofactor, it is meant as a reference to any protein having an amino

acid sequence substantially the same as SEQ ID NO. 3 for CF16, SEQ ID NO. 6 for CF17, SEQ ID NO. 9 for CF18, SEQ ID NO. 12 for CF19, SEQ ID NO. 15 for CF40, SEQ ID NO. 18 for CF41, SEQ ID NO. 21 for CF42, and SEQ ID NO. 24 for CF43.

The present invention also comprises the nucleic acid sequences encoding the cofactors 16 to 20, which nucleic acid sequences are substantially the same as SEQ ID NO. 1 for CF16, SEQ ID NO. 4 for CF17, SEQ ID NO. 7 for CF18, SEQ ID NO. 10 for CF19, SEQ ID NO. 13 for CF40, SEQ ID NO. 16 for CF41, SEQ ID NO. 19 for CF42, and SEQ ID NO. 22 for CF43 all encoding human cofactors as preferred embodiments and/or the complements thereof as shown in SEQ ID NO. 2 for CF16, SEQ ID NO. 5 for CF17, SEQ ID NO. 8 for CF18, SEQ ID NO. 11 for CF19, SEQ ID NO. 14 for CF40, SEQ ID NO. 17 for CF41, SEQ ID NO. 20 for CF42, and SEQ ID NO. 23 for CF43.

Herein the "complement" refers to the complementary strand of the nucleic acid according to the invention, thus the strand that would hybridize to the nucleic acid according to the invention. In accordance with standard biological terminology all DNA sequences herein are however written in 5'-3' orientation, thus the complements depicted are actually "reverse" complements For simplification purposes they are however some times referred to simply as "complements".

As used herein, a protein "having an amino acid sequence substantially the same as SEQ ID NO x" (where "x" is the number of one of the protein sequences recited in the Sequence Listing) means a protein whose amino acid sequence is the same as SEQ ID NO x or differs only in a way such that at least 50% of the residues compared in a sequence alignment with SEQ ID NO. x are identical, preferably 75% of the residues are identical, even more preferably 95% of the residues are identical and most preferably at least 98% of the residues are identical

Those skilled in the art will appreciate that conservative substitutions of amino acids can be made without significantly diminishing the protein's affinity for interacting proteins, DNA binding sites, cofactor modulators, e.g. small molecular hydrophobic compounds, or RNA.

Other substitutions may be made that increase the proteins' affinity for these compounds. Making and identifying such proteins is a routine matter given the teachings herein, and can be accomplished, for example, by altering the nucleic acid sequence encoding the protein (as

disclosed herein), inserting it into a vector, transforming a cell, expressing the nucleic acid sequence, and measuring the binding affinity of the resulting protein, all as taught herein.

As used herein the term "a molecule having a nucleotide sequence substantially the same as SEQ ID NO y" (wherein "y" is the number of one of the protein-encoding nucleotide sequences listed in the Sequence Listing) means a nucleic acid encoding a protein "having an amino acid sequence substantially the same as SEQ ID NO y+1" (wherein "y+1" is the number of the amino acid sequence for which nucleotide sequence "y" codes) as defined above. This definition is intended to encompass natural allelic variations in the CF sequences. Cloned nucleic acid provided by the present invention may encode CF proteins of any species of origin, including (but not limited to), for example, mouse, rat, rabbit, hamster, cat, dog, pig, primate, and human.

Preferably the nucleic acids provided by the invention encode CFs of mammalian, preferably mouse and most preferably human origin.

IDENTIFICATION OF VARIANTS AND HOMOLOGUES AS WELL AS USE OF PROBES:

Nucleic acid hybridization probes provided by the invention are nucleic acids consisting essentially of the nucleotide sequences complementary to any sequence depicted in SEQ ID NO. 1, 4, 7, 10, 13, 16, 19, and 22, and/or the complements thereof as shown in SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, and 23, or parts thereof which are effective in nucleic acid hybridization.

Nucleic acid hybridization probes provided by the invention are nucleic acids capable of detecting *i.e.* hybridizing to the gene encoding the polypeptides according to SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, and 24.

Nucleic acid probes are useful for detecting CF gene expression in cells and tissues using techniques well-known in the art, including, but not limited to, Northern blot hybridization, in situ hybridization, and Southern hybridization to reverse transcriptase - polymerase chain reaction product DNAs. The probes provided by the present invention, including oligonucleotide probes derived therefrom, are also useful for Southern hybridization of mammalian, preferably human, genomic DNA for screening for restriction fragment length polymorphism

(RFLP) associated with certain genetic disorders. As used herein, the term complementary means a nucleic acid having a sequence that is sufficiently complementary in the Watson-Crick sense to a target nucleic acid to bind to the target under physiological conditions or experimental conditions those skilled in the art routinely use when employing probes.

It is understood in the art that a nucleic acid sequence will hybridize with a complementary nucleic acid sequence under highly stringent conditions as defined herein, even though some mismatches may be present. Such closely matched, but not perfectly complementary sequences are also encompassed by the present invention. For example, differences may occur through genetic code degeneracy, or by naturally occurring or man made mutations and such mismatched sequences would still be encompassed by the present claimed invention.

Preferably, the nucleotide sequences of the nuclear cofactors SEQ ID NOs: 1, 4, 7, 10, 13, 16, 19, and 22, and/or their complements SEQ ID NOs 2, 5, 8, 11, 14, 17, 20, and 23 can be used to derive oligonucleotide fragments (probes) of various length. Stretches of 17 to 30 nucleotides are used frequently but depending on the screening parameters longer sequences as 40, 50, 100, 150 up to the full length of the sequence may be used. Those probes can be synthesized chemically and are obtained readily from commercial oligonucleotide providers. Chemical synthesis has improved over the years and chemical synthesis of oligonucleotides as long as 100-200 bases is possible. The field might advance further to allow chemical synthesis of even longer fragments. Alternatively, probes can also be obtained by biochemical *de novo* synthesis of single stranded DNA. In this case the nucleotide sequence of the nuclear receptors or their complements serve as a template and the corresponding complementary strand is synthesized. A variety of standard techniques such as nick translation or primer extension from specific primers or short random oligonucleotides can be used to synthesize the probe (Sambrook, J., Fritsch, E.F. & Maniatis , T. Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, 1989)). Nucleic acid reproduction technologies exemplified by the polymerase chain reaction (Saiki, R.K. et *al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491 (1988)) are commonly applied to synthesize probes. In the case of techniques using specific primers the nucleic acid sequences of the nuclear receptors or their complements are not only used as a template in the biochemical reaction but also to derive the specific primers which are needed to prime the reaction.

In some cases one might also consider to use the nucleic acid sequence of the cofactors or their complements as a template to synthesize an RNA probe. A promoter sequence for a DNA-dependent RNA polymerase has to be introduced at the 5'-end of sequence. As an example this can be done by cloning the sequence in a vector which carries the respective promoter sequence. It is also possible to introduce the needed sequence by synthesizing a primer with the needed promoter in the form of a 5' "tail". The chemical synthesis of a RNA probe is another option.

Appropriate means are available to detect the event of a hybridization. There is a wide variety of labels and detection systems, e.g. radioactive isotopes, fluorescent, or chemiluminescent molecules which can be linked to the probe. Furthermore, there are methods of introducing haptens which can be detected by antibodies or other ligands such as the avidin/biotin high affinity binding system.

Hybridization can take place in solution or on solid phase or in combinations of the two, e.g. hybridization in solution and subsequent capture of the hybridization product onto a solid phase by immobilized antibodies or by ligand coated magnetic beads.

Hybridization probes act by forming selectively duplex molecules with complementary stretches of a sequence of a gene or a cDNA. The selectivity of the process can be controlled by varying the conditions of hybridization. To select sequences which are identical or highly homologous to the sequence of interest stringent conditions for the hybridization will be used, e.g. low salt in the range of 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade. Stringency can be further improved by the addition of formamide to the hybridisation solution. The use of stringent conditions which means that only little mismatch or a complete match will lead to a hybridization product would be used to isolate closely related members of the same gene family. Thus, as used herein stringent hybridization conditions are those where between 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade are applied.

The use of highly stringent conditions or conditions of "high stringency" means that only very little mismatch or a complete match which lead to a hybridization product would be used to isolate very closely related members of the same gene family. Thus, as used herein highly

stringent hybridization conditions are those where between 0.02 – 0.3 M salt and 65°C degrees centigrade are applied for about 5 to 18 hours of hybridization time and additionally, the sample filters are washed twice for about 15 minutes each at between 60°C – 65°C degrees centigrade, wherein the first washing fluid contains about 0.1 M salt (NaCl and/or Sodium Citrate) and the second contains only about 0.02 M salt (NaCl and/or Sodium Citrate). In a preferred embodiment the following conditions are considered to be highly stringent:

Hybridisation in a buffer containing 2 x SSC (0.03 M Sodium Citrate, 0.3 M NaCl) at 65°C – 68°C degrees centigrade for 12 hours, followed by a washing step for 15 minutes in 0.5 x SSC, 0.1% SDS, and a washing step for 15 minutes at 65°C degrees centigrade in 0.1 x SSC, 0.1% SDS.

Less stringent hybridization conditions, e.g. 0.15 M salt - 1 M salt and/or temperatures from 22°C degrees centigrade to 56°C degrees centigrade are applied in order to detect functionally equivalent genes in the same species or for orthologous sequences from other species.

Unspecific hybridization products are removed by washing the reaction products repeatedly in 2 x SSC solution and increasing the temperature.

DEGENERATE PCR AND CLONING OF HOMOLOGUES

The nucleotide sequences of the cofactors CF16 to CF19 and CF40 to CF43 or their complements can be used to design primers for a polymerase chain reaction. Due to the degeneracy of the genetic code the respective amino acid sequence is used to design oligonucleotides in which varying bases coding for the same amino acid are included. Numerous design rules for degenerate primers have been published (Compton et al, 1990). As in hybridization there are a number of factors known to vary the stringency of the PCR. The most important parameter is the annealing temperature. To allow annealing of primers with imperfect matches annealing temperatures are often much lower than the standard annealing temperature of 55°C, e.g. 35°C to 52°C degrees can be chosen. PCR reaction products can be cloned. Either the PCR product is cloned directly, with reagents and protocols from commercial manufacturers (*e.g.* from Invitrogen, San Diego, USA). Alternatively, restriction sites can be introduced intro the PCR product via a 5'-tail of the PCR primers and used for cloning.

GENETIC VARIANTS

Fragments from the nucleotide sequence of the cofactors or their complements can be used to cover the whole sequence with overlapping sets of PCR primers. These primers are used to produce PCR products using genomic DNA from a human diversity panel of healthy individuals or genomic DNA from individuals which are phenotypically conspicuous. The PCR products can be screened for polymorphisms, for example by denaturing gradient gel electrophoresis, binding to proteins detecting mismatches or cleaving heteroduplices or by denaturing high-performance liquid chromatography. Products which display mutations need to be sequenced to identify the nature of the mutation. Alternatively, PCR products can be sequenced directly omitting the mutation screening step to identify genetic polymorphisms. If genetic variants are identified and are associated with a discrete phenotype, these genetic variations can be included in diagnostic assays. The normal variation of the human population is of interest in designing screening assays as some variants might interact better or worse with a respective lead, i.e. therapeutic or potentially therapeutic substance (a pharmacodynamic application). Polymorphisms or mutations which can be correlated to phenotypic outcome are a tool to extend the knowledge and the commercial applicability of the nucleotide sequences of the cofactors CF16 to CF19 and CF40 to CF43 or their complements or their gene products, as variants might have a slightly different molecular behavior or desired properties. Disease-causing mutations or polymorphisms allow the replacement of this disease inducing gene copy with a wild-type copy by means of gene therapy approaches and/or the modulation of the activity of the gene product by drugs.

PREPARATION OF POLYNUCLEOTIDES:

DNA which encodes cofactor CF16 to CF19 and CF40 to CF43 may be obtained, in view of the instant disclosure, by chemical synthesis, by screening reverse transcripts of mRNA from appropriate cells or cell line cultures, by screening genomic libraries from appropriate cells, or by combinations of these procedures, as illustrated below.

Screening of mRNA or genomic DNA may be carried out with oligonucleotide probes generated from the CF nucleotide sequences information provided herein. These oligonucleotides are in addition useful to isolate a full length cDNA from an appropriate cDNA library.

Probes may be labeled with a detectable group such as a fluorescent group, a radioactive atom or a chemiluminescent group in accordance with known procedures and used in conventional hybridization assays, as described in greater detail in the examples below. Alternatively, the CF nucleotide sequences may be obtained by use of the polymerase chain reaction (PCR) procedure, with the PCR oligonucleotide primers being produced from the CF nucleotide sequences provided herein, according to SEQ ID NO 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, and SEQ ID NO. 22, and/or the complements thereof as shown in SEQ ID NO. 2, SEQ ID NO. 5, SEQ ID NO. 8, SEQ ID NO. 11, SEQ ID NO. 14, SEQ ID NO. 17, SEQ ID NO. 20, and SEQ ID NO. 23, or parts thereof.

Upon purification or synthesis, the nucleic acid according to the invention may be labeled, e.g. for use as a probe.

As single and differential labeling agents and methods, any agents and methods which are known in the art can be used provided that they do not significantly altering the stability or function of said primer in the DNA sequencing method of the present invention. For example, single and differential labels may consist of the group comprising enzymes such as β-galactosidase, alkaline phosphatase and peroxidase, enzyme substrates, coenzymes, dyes, chromophores, fluorescent, chemiluminescent and bioluminescent labels such as FITC, Cy5, Cy5.5, Cy7, Texas-Red and IRD40(Chen et al. (1993), J. Chromatog. A 652: 355-360 and Kambara et al. (1992), Electrophoresis 13: 542-546), ligands or haptens such as biotin, and radioactive isotopes such as ^{3}H, ^{35}S, ^{32}P ^{125}I and ^{14}C.

EXPRESSION OF THE CF16 TO CF19 AND CF40 TO CF43 PROTEINS/POLYPETIDES:

The CF nucleic acids or polypeptides may be synthesized in host cells transformed with a recombinant expression construct comprising a nucleic acid encoding any of the cofactors according to the invention, namely CF16 to CF19 and CF40 to CF43. Such a recombinant expression construct can also be comprised of a vector that is a replicable DNA construct.

Amplification vectors do not require expression control domains. All that is needed is the ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants. See, Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd Edition, Cold Spring Harbor Press, New York, 1989).

An expression vector comprises a polynucleotide operatively linked to a prokaryotic promoter. Alternatively, an expression vector is a polynucleotide operatively linked to an enhancer-promoter that is a eukaryotic promoter, and the expression vector further has a polyadenylation signal that is positioned 3' of the carboxy-terminal amino acid and within a transcriptional unit of the encoded polypeptide. A promoter is a region of a DNA molecule typically within about 500 nucleotide pairs in front of (upstream of) the point at which transcription begins (*i.e.*, a transcription start site). In general, a vector contains a replicon and control sequences which are derived from species compatible with the host cell. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells.

Another type of discrete transcription regulatory sequence element is an enhancer. An enhancer provides specificity of time, location and expression level for a particular encoding region (e.g., gene). A major function of an enhancer is to increase the level of transcription of a coding sequence in a cell.

As used herein, the phrase "enhancer-promoter" means a composite unit that contains both enhancer and promoter elements. An enhancer-promoter is operatively linked to a coding sequence that encodes at least one gene product.

An enhancer-promoter used in a vector construct of the present invention may be any enhancer-promoter that drives expression in a prokaryotic or eukaryotic cell to be transformed/transfected.

A coding sequence of an expression vector is operatively linked to a transcription terminating region. RNA polymerase transcribes an encoding DNA sequence through a site where polyadenylation occurs.

An expression vector that comprises a polynucleotide that encodes one of the the polypeptides of the cofactors CF16 to CF19 and CF40 to CF43 is meant to include a sequence of nucleotides encoding a CF polypeptide sufficient in length to distinguish said segment from a polynucleotide segment encoding a non- cofactor polypeptide.

A polypeptide of the invention may also encode biologically functional polypeptides or peptides which have variant amino acid sequences, such as with changes selected based on considerations such as the relative hydropathic score of the amino acids being exchanged.

These variant sequences are those isolated from natural sources or induced in the sequences disclosed herein using a mutagenic procedure such as site-directed mutagenesis.

Furthermore, an expression vector of the present invention may contain regulatory elements for optimized translation of the polypeptide in prokaryotic or eukaryotic systems. These sequences are operatively located around the transcription start site and are most likely similar to ribosome recognition sites like prokaryotic ribosome binding sites (RBS) or eukaryotic Kozak sequences as known in the art (Kozak M., Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187-208 (1999)).

An expression vector of the present invention is useful both as a means for preparing quantities of the CFs' polypeptide-encoding DNA itself, and as a means for preparing the encoded CFs' polypeptide and peptides. It is contemplated that where cofactor polypeptides of the invention are made by recombinant means, one may employ either prokaryotic or eukaryotic expression vectors as shuttle systems.

Where expression of recombinant CF16 to CF19 and CF40 to CF43 polypeptide is desired and a eukaryotic host is contemplated, it is most desirable to employ a vector such as a plasmid, that incorporates a eukaryotic origin of replication. Additionally, for the purposes of expression in eukaryotic systems, one desires to position the cofactor encoding sequence or if desired parts thereof adjacent to and under the control of an effective eukaryotic promoter. To bring a coding sequence under control of a promoter, whether it is eukaryotic or prokaryotic, what is generally needed is to position the 5' end of the translation initiation side of the proper translational reading frame of the polypeptide between about 1 and about 2000 nucleotides 3' of or downstream with respect to the promoter chosen.

Furthermore, where eukaryotic expression is anticipated, one would typically desire to incorporate into the transcriptional unit which includes the CF polypeptide, an appropriate polyadenylation side.

The invention provides homogeneous compositions of mammalian cofactor polypeptides produced by transformed prokaryotic or eukaryotic cells as provided herein. Such homogeneous compositions are intended to be comprised of mammalian cofactor protein that comprises at least 90% of the protein in such homogenous composition. The invention also provides membrane preparation from cells expressing the mammalian cofactors polypeptides as the result of transformation with a recombinant expression construct, as described here.

Within the scope of the present invention the terms recombinant protein or coding sequence both also include tagged versions of the polypeptides depicted in SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12, SEQ ID NO. 15, SEQ ID NO. 18, SEQ ID NO. 21, and/or SEQ ID NO. 24 and fusion proteins of said proteins with any other recombinant protein. Tagged versions here means that small epitopes of 3-20 amino acids are added to the original protein by extending the coding sequence either at the 5´or the 3´terminus leading to N-terminal or C-terminal extended proteins respectively, or that such small epitopes are included elsewhere in the protein. The same applies for fusion proteins where the added sequences are coding for longer proteins, varying between 2 and 100 kDa. Tags and fusion proteins are usually used to facilitate purification of recombinant proteins by specific antibodies or affinity matrices or to increase solubility of recombinant proteins within the expression host. Fusion proteins are also of major use as essential parts of yeast two hybrid screens for interaction partners of recombinant proteins.

Tags used in the scope of the present invention may include but are not limited to the following: EEF (alpha Tubulin), B-tag (QYPALT), E tag (GAPVPYPDPLEPR) c-myc Tag (EQKLISEEDL), Flag epitope (DYKDDDDK), HA tag (YPYDVPDYA), 6 or 10 x His Tag, HSV (QPELAPEDPED), Pk-Tag (GKPIPNPLLGLDST), protein C (EDQVDPRLIDGK), T7 (MASMTGGQQMG), VSV-G (YTDIEMNRLGK), Fusion proteines may include Thioredoxin, Glutathiontransferase (GST), Maltose binding Protein (MBP), Cellulose Binding protein, calmodulin binding protein, chitin binding protein, ubiquitin, the Fc part of Immunoglobulins, and the IgG binding domain of Staphylococcus aureus protein A. These examples of course are illustrative and not limiting.

For expression of recombinant proteins in living cells or organisms, vector constructs harboring recombinant cofactors as set forth in SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, and/or SEQ ID NO. 22 and/or the complements thereof SEQ ID NO. 2, SEQ ID NO. 5, SEQ ID NO. 8, SEQ ID NO.

11, SEQ ID NO. 14, SEQ ID NO. 17, SEQ ID NO. 20, and/or SEQ ID NO. 23 are transformed or transfected into appropriate host cells. Preferably, a recombinant host cell of the present invention is transfected with a polynucleotide SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, and/or SEQ ID NO. 22.

Means of transforming or transfecting cells with exogenous polynucleotide such as DNA molecules are well known in the art and include techniques such as calcium-phosphate- or DEAE-dextran-mediated transfection, protoplast fusion, electroporation, liposome mediated transfection, direct microinjection and virus infection (Sambrook et al., 1989).

The most frequently applied technique for transformation of prokaryotic cells is transformation of bacterial cells after treatment with calcium chloride to increase permeability (Dagert & Ehrlich, 1979), but a variety of other methods is also available for one skilled in the art.

The most widely used method for transfection of eukaryotic cells is transfection mediated by either calcium phosphate or DEAE-dextran. Although the mechanism remains obscure, it is believed that the transfected DNA enters the cytoplasm of the cell by endocytosis and is transported to the nucleus. Depending on the cell type, up to 90% of a population of cultured cells may be transfected at any one time. Because of its high efficiency, transfection mediated by calcium phosphate or DEAE-dextran is the method of choice for studies requiring transient expression of the foreign nucleic acid in large numbers of cells. Calcium phosphate-mediated transfection is also used to establish cell lines that integrate copies of the foreign DNA, which are usually arranged in head-to-tail tandem arrays into the host cell genome.

In the protoplast fusion method, protoplasts derived from bacteria carrying high numbers of copies of a plasmid of interest are mixed directly with cultured mammalian cells. After fusion of the cell membranes (usually with polyethylene glycol), the contents of the bacterium are delivered into the cytoplasm of the mammalian cells and the plasmid DNA is transported to the nucleus. Protoplast fusion is not as efficient as transfection for many of the cell lines that are commonly used for transient expression assays, but it is useful for cell lines in which endocytosis of DNA occurs inefficiently. Protoplast fusion frequently yields multiple copies of the plasmid DNA tandemly integrated into the host chromosome.

The application of brief, high-voltage electric pulses to a variety of mammalian and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation may be extremely efficient and may be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

Liposome transfection involves encapsulation of DNA and RNA within liposomes, followed by fusion of the liposomes with the cell membrane. The mechanism of how DNA is delivered into the cell is unclear but transfection efficiencies may be as high as 90%.

Direct microinjection of a DNA molecule into nuclei has the advantage of not exposing DNA to cellular compartments such as low-pH endosomes. Microinjection is therefore used primarily as a method to establish lines of cells that carry integrated copies of the DNA of interest.

The use of adenovirus as a vector for cell transfection is well known in the art. Adenovirus vector-mediated cell transfection has been reported for various cells (Stratford-Perricaudet et al., 1992).

Furthermore, the possibility exists, to perform the gene transfer *in vivo*, either by preferential stereotactic injection of the infectious particle or by direct application of virus-producing cells (Oldfield, et al. Hum. Gen. Ther., 1993, 4:39-69).

The commonly used viral vectors for the transfer of genes according to the current state of the art are mainly retroviral, lentiviral, adenoviral and adeno-associated viral vectors. These are circular nucleotide sequences derived from natural viruses in which at least the viral structural protein encoding genes are replaced by the construct to be transferred.

Retroviral vector systems provide the prerequisite for a long-lasting expression of the transgene by the stable, but non-directed integration into the genome of the host. Vectors of the younger generation possess no irrelevant and potentially immunogenic proteins, furthermore,

there is no pre-existing immunity of the recipient in view of the vector. Retroviruses contain an RNA-genome, that is packed into a lipid coating, which consists out of parts of the host cell membrane and viral proteins. For the expression of viral genes the RNA-genome is reversely transcribed and integrated into the target-cell DNA using the enzyme integrase. This can subsequently be transcribed and translated by the infected cell, thereby viral compounds are produced which then form retrovirus particles. RNA will then be exclusively included in the newly produced viruses. The genome of retroviruses contains three essential genes: *gag*, which codes for viral structural proteins, so-called group-specific antigenes, *pol* for Enzyme like reverse transcriptase and integrase and *env* for the „envelope" protein, which is responsible for the binding of the host specific receptor. The production of the replication incompetent viruses occurs after transfection in so-called packaging-celllines which are in addition provided with gag/pol-encoding genes and express those „in trans" and thereby complement the formation of replicationincompeten (i.e. gag/pol-deletet) transgene virus particles. An alternative is cotransfection of the essential virus genes, wherein only the transgen containing vector carries the packaging signal.

Novel, non-viral vectors consist out of autonomically and self-integrating DNA sequences, the transposons, which are introduced into the host cell by, e.g. liposomal transfection and were for the first time successfully used for the expression of human transgenes in mammalian cells (Yant et al., 2000).

A transfected cell may be prokaryotic or eukaryotic, transfection may be transient or stable. Where it is of interest to produce a full length human CF16 to CF19 and CF40 to CF 43 protein, cultured mammalian or human cells are of particular interest.

In another aspect, the recombinant host cells of the present invention are prokaryotic host cells. In addition to prokaryotes, eukaryotic microbes, such as yeast may also be used illustrative examples for suitable cells and organisms for expression of recombinant proteins are belonging to but not limited to the following examples: Insect cells, such as Drosophila Sf21, SF9 cells or others, Expression strains of Escherichia coli, such as XL1 blue, BRL21, M15, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Hansenula polymorpha and Pichia pastoris strains, immortalized mammalian cell lines such as AtT-20, VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COSM6, COS-7, 293 and MDCK cells, BHK-21 cells, Att 20HeLa cells, HeK 294, T47 D cells and others.

Expression of recombinant proteins within the scope of this invention can also be performed in vitro. This may occur by a two step procedure, thereby producing first mRNA by in vitro transcription of an apt polynucleotide construct followed by in vitro translation with convenient cellular extracts. These cellular extracts may be reticulocyte lysates but are not limited to this type. In vitro transcription may be performed by T7 or SP6 DNA polymerase or any other RNA polymerase which can recognize per se or with the help of accessory factors the promoter sequence contained in the recombinant DNA construct of choice. Alternatively one of the recently made available one step coupled transcription/translation systems may be used for in vitro translation of DNA coding for the proteins of this invention. One illustrative but not limiting example for such a system is the TNT® T7 Quick System by Promega.

Expression of recombinant proteins in transfected cells may occur constitutively or upon induction. Procedures depend on the Cell/vector combination used and are well known in the art.

In all cases, transfected cells are maintained for a period of time sufficient for expression of the recombinant cofactor proteins according to the invention. A suitable maintenance time depends strongly on the cell type and organism used and is easily ascertainable by one skilled in the art. Typically, maintenance time is from about 2 hours to about 14 days. For the same reasons and for sake of protein stability and solubility incubation temperatures during maintenance time may vary from 20°C to 42 °C.

Recombinant proteins are recovered or collected either from the transfected cells or the medium in which those cells are cultured. Recovery comprises cell disruption, isolation and purification of the recombinant protein. Isolation and purification techniques for polypeptides are well-known in the art and include such procedures as precipitation, filtration, chromatography, electrophoresis and the like.

In a preferred embodiment, purification includes but is not limited to affinity purification of tagged or nontagged recombinant proteins. This is a well established robust technique easily adapted to any tagged protein by one skilled in the art. For affinity purification of tagged proteines, small molecules such as gluthathione, maltose or chitin, specific proteins such as the IgG binding domain of Staphylococcus aureus protein A, antibodies or specific chelates

which bind with high affinity to the tag of the recombinant protein are employed. For affinity purification of non-tagged proteins specific monoclonal or polyclonal antibodies, which were raised against said protein, can be used. Alternatively immobilized specific interactors of said protein may be employed for affinity purification. Interactors include native or recombinant proteins as well as native or artificial specific low molecular weight ligands.

CHEMICAL SYNTHESIS OF THE POLYPEPTIDES ACCORDING TO THE INVENTION:

Alternatively, the protein itself may be produced using chemical methods to synthesize any of the amino acid sequences according to the invention (SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12, SEQ ID NO. 15, SEQ ID NO. 18, SEQ ID NO. 21, and/or SEQ ID NO. 24) or that is encoded by the nucleotide sequences according to the invention and/or the complements thereof or a portion thereof. For example, peptide synthesis can be performed using conventional Merrifield solid phase f-Moc or t-Boc chemistry or various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269: 202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer). The newly synthesized peptide(s) may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.,* the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequences according to the invention, *i.e.* SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12, SEQ ID NO. 15, SEQ ID NO. 18, SEQ ID NO. 21, and SEQ ID NO. 24 or the sequence that is encoded by SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, and SEQ ID NO. 22 or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

COMPLEXES OF THE COFACTORS ACCORDING TO THE INVENTION WITH OTHER POLYPETIDES

As outlined above CF16 to CF19 and CF40 to CF43 all bind ER alpha, presumably also *in vivo*. In a preferred embodiment of the invention one or more of the CFs are complexed with the ER alpha polypeptide or portions thereof, preferentially *in vitro*. The ER alpha polypep-

tide receptor is encoded by a genomic nucleic acid sequence according to SEQ ID NO. 25 or 38. The receptor has an amino acid sequence according to SEQ ID NO. 27 or 30.

Such complexes are particularly suited for all forms of binding or screening assays (see also below). Thus, in a preferred embodiment of the invention such assays are performed with complexes of the receptor(s) associated with one or more of the CF proteins.

In one embodiment of the invention a trimeric complex is claimed consisting of ER alpha homodimers bound to one of the CFs. In another embodiment of the invention ER alpha may bind in monomeric form to one of the CFs. Such complexes may be used in binding and screening assays as outlined below.

In one embodiment of the invention the entire CF polypeptide is part of the complex or alternatively only a portion, e.g. a truncated fragment of the other polypeptide (ER alpha) is part of the complex.

SCREENING ASSAYS

In still a further embodiment, the present invention concerns a method for identifying new inhibitory or stimulatory substances of the cofactors according to the invention, these substances may be termed as "candidate substances". It is contemplated that this screening technique proves useful in the general identification of compounds that serve the purpose of inhibiting or stimulating cofactor activity.

The following substances are interactors of the ER alpha-cofactor complex according to the invention:

tamoxifen
4-hydroxytamoxifen
Deaminohydroxy)toremifene (Z-2-[4-(4-chloro- 1,2-diphenyl-but-1-enyl)phenoxy]ethanol; FC-1271a
idoxifene
raloxifene (LY139481 HCI)
genistein
toremifene
ICI 182,780 (Faslodex)
coumestrol
yuehchukene
estrogen (17beta-estradiol;E2)

7-ketocholestanol
5 alpha-androstane-3 beta, 17 beta-diol
3 beta GSD (gestodene)
3 beta-Hydroxy-5alpha-androstan-17-one (DHEA)
bisphenol A
estriol
estrone
16 alpha-Hydroxyestrone (16OHE1)
11 beta-chloromethyl-estradiol-17 beta
diethyl-stilbestrol
hexestrol
clomiphene
hydroxylated triphenylacrylonitriles
17 alpha-estradiol
chlordecone (Kepone)
pyrrolo[2,1,5-cd] indolizine (NNC 45-0095)
FC1271a (triphenylethylene compound)
forskolin
diethylstilbestrol-4',4"-quinone
CP-336,156
panomifene (EGIS-5660)
2,2-bis-(p-hydroxyphenyl)-1,1,1-trichloroethane (HPTE)
toremifene
bis(4-hydroxyphenyl)[2-(phenoxysulfonyl)phenyl]methane
triphenylethylene H1285
17 alpha-[125I]iodovinyl-11 beta-methoxyestradiol
RU 16117
desmethyltamoxifen
dichlorodiphenyltrichloroethane (DDT)
11beta-substituted 21-chloro/iodo-(17alpha,20E/Z)-19-norpregna-1,3,5(10),20-te
EM-800

Accordingly, in screening assays for identifying pharmaceutical agents which affect cofactor acitivity, it is proposed that compounds isolated from natural sources, such as fungal extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents.

It will be understood that that the pharmaceutical agents to be screened can also be derived from chemical compositions or man-made compounds. The candidate substances could also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are other-

wise inactive. We anticipate that such screens will in some cases lead to the isolation of agonists of nuclear receptors or cofactors, in other cases to the isolation of antagonists. In other instances, substances will be identified that have mixed agonistic and antagonistic effects, or affect nuclear receptors or cofactors in any other way.

In another embodiment, the invention concerns the isolation of substance inhibiting the interaction of the cofactor protein and ER alpha. Such substances are useful for the development of drugs against diseases as listed above. Substances disrupting the interactions may be isolated by a variety of screening methods including the two hybrid system or the reverse two hybrid system (Lenna C.A. and Hannink, M. 1996, Nucl. Acids Res. 24: 3341-3347), or any variation of cellular or cell free assays as described in this invention, as is obvious to anyone skilled in the art.

In an important embodiment of the invention, the binding of the cofactor protein and ER alpha can be used to monitor the binding of a substance to one of the binding partners. The substance, which can be a small molecule such as a ligand to a nuclear receptor, will lead to a change in the allosteric conformation of the binding protein which in consequence leads to a loss of the interaction of the two proteins. Using this effect of ligand-dependent protein-protein interactions one can design assays where the protein-protein interaction serves as a surrogate read-out for the binding of one of the proteins to small molecule ligand. Any assay method which is useful for the measurement of protein-protein interactions can be used for such an indirect assay. Such assay methods are well known in the art and include the methods described in this patent under "Cell free assays" and "Cell based assays". In a preferre d embodiment, this assay will measure the binding of substances to ER alpha, resulting in an effect on the interaction of ER alpha with the cofactor.

CELL BASED ASSAYS

To identify a candidate substance capable of influencing the cofactor protein activity, one first obtains a recombinant cell line. One designs the cell line in such a way that the activity of the cofactor leads to the expression of a protein which has an easily detectable phenotype (a reporter), such as luciferase, fluorescent proteins such as green or red fluorescent protein, beta-galactosidase, alpha-galactosidase, beta-lactamase, chloramphenicol-acetyl-transferase, beta-glucuronidase, or any protein which can be detected by a secondary reagent such as an antibody.

Methods for detecting proteins using antibodies, such as ELISA assays, are well known to those skilled in the art.

Here, the amount of reporter protein present reflects the activity of the cofactor. This recombinant cell line is then screened for the effect of substances on the expression of the reporters, thus measuring the effect of these substances on the activity of the cofactor. These substances can be derived from natural sources, such as fungal extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents. It will be understood that that the pharmaceutical agents to be screened may be derived from chemical compositions or man-made compounds.

The candidate substances can also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are otherwise inactive.

In general the assay can be performed by firstly bringing a suitable cell containing a reporter gene which transcription is influenced by the cofactors activity in contact with a compound and secondly monitoring the expression of the reporter gene to evaluate the effect of the compound on the activity of the cofactor.

In other embodiments of the invention assays are included where measuring the activity of di- or multimeric complexes of the cofactor and other proteins such as ER alpha. Further included are assays aiming at the identification of compounds which specifically influence only the monomeric, homodimeric or homomultimeric form of the cofactor, or influencing only multimeric forms of the cofactor. Such assays include measuring the effect of a compound on the cofactor in the absence of a binding partner, and measuring the effect of a compound on the cofactor in the presence of a binding partner, such as ER alpha. One skilled in the art will find numerous more assays which are equally covered by the invention.

A cell line where the activity of ER alpha or any other nuclear receptor determines the expression of a reporter can be obtained by generating an artificial promoter upstream of the reporter gene, which contains preferably multiple copies of HREs to which ER alpha or any other nuclear receptor binds.

Furthermore, transgenic animals described in the invention can be used to derive cell lines useful for cellular screening assays.

Cell lines useful for such an assay include many different kinds of cells, including prokaryotic, animal, fungal, plant and human cells. Yeast cells can be used in this assay, including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* cells.

One way of building cellular assays is by measuring the effect of compounds is the use of the two hybrid system (see for example, U.S. Pat. No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; PCT Publication No. WO 94/10300, and U.S. Pat. No. 5,667,973), or possible variants of the basic two hybrid system as discussed e.g in Vidal M, Legrain P, Nucleic Acids Res 1999 Feb 15;27(4):919-29. Briefly, the two hybrid assay relies on reconstituting in vivo a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a cofactor. The second hybrid protein encodes a transcriptional activation domain fused in frame to another gene, for example ER alpha. If the cofactor and ER alpha proteins are able to interact, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the cofactor and ER alpha proteins. Suitable host cells for such assays include yeast cells, but also mammalian cells or bacterial cells.

In such assays, one primarily measures the effect of a compound on a given interaction involving one of the CF16 to CF19 and CF40 to CF43 cofactors and a binding protein. In a pre-

ferred embodiment of the invention systems using other hosts such as prokaryotes as *E. coli*, or eukaryotic mammalian cells are described.

Two hybrid systems using hybrid protein fusions with other proteins than transcription factors, including enzymes such as beta-galactosidase or *dihydrofolate reductase* may also be applied. These assays are useful both to monitor the effect of a compound, including peptides, proteins or nucleic acids on an interaction of a cofactor with a given binding partner, as well as to identify novel proteins or nucleic acids interacting with the cofactor.

CELL-FREE ASSAYS

Recombinant forms of the polypeptides according to SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12, SEQ ID NO. 15, SEQ ID NO. 18, SEQ ID NO. 21, or SEQ ID NO. 24 can be used in cell-free screening assays aiming at the isolation of compounds affecting the activity of cofactors. In such an assay, the cofactor polypeptides are brought into contact with a substance to test if the substance has an effect on the activity of the cofactor.

The detection of an interaction between an agent and a cofactor may be accomplished through techniques well-known in the art. These techniques include but are not limited to centrifugation, chromatography, electrophoresis and spectroscopy. The use of isotopically labeled reagents in conjunction with these techniques or alone is also contemplated. Commonly used radioactive isotopes include ^{3}H, ^{14}C, ^{22}Na, ^{32}P, ^{33}P, ^{35}S, ^{45}Ca, ^{60}Co, ^{125}I, and ^{131}I. Commonly used stable isotopes include ^{2}H, ^{13}C, ^{15}N, ^{18}O.

For example, if an agent binds to any of the cofactors of the present invention, the binding may be detected by using radiolabeled agent or radiolabeled cofactor. Briefly, if radiolabeled agent or radiolabeled cofactor is utilized, the agent-cofactor complex may be detected by liquid scintillation or by exposure to x-ray film or phosho-imaging devices.

One way to screen for substances affecting cofactor activity is to measure the effect of the substance on the binding affinity of the cofactor to other proteins or molecules, such as activators or repressors, DNA, RNA, other proteins, antibodies peptides or other substances, including chemical compounds known to affect receptor activity or to a nuclear receptor itself. Assays measuring the binding of a protein to a ligand are well known in the art, such as

ELISA assays, FRET assays, bandshift assays, plasmon-resonance based assays, scintilllation proximity assays, fluorescence polarization assays, alpha screen assays.

In one example, a mixture containing a cofactor polypeptide, effector and candidate substance is allowed to incubate. The unbound effector is separable from any effector/cofactor complex so formed. One then simply measures the amount of each (*e.g.*, versus a control to which no candidate substance has been added). This measurement may be made at various time points where velocity data is desired. From this, one determines the ability of the candidate substance to alter or modify the function of the cofactor.

Numerous techniques are known for separating the effector from effector/cofactor complex, and all such methods are intended to fall within the scope of the invention. This includes the use of thin layer chromatographic methods (TLC), HPLC, spectrophotometric, gas chromatographic/mass spectrophotometric or NMR analyses. Another method of separation is to immobilize one of the binding partners on a solid support, and to wash away any unbound material. It is contemplated that any such technique may be employed so long as it is capable of differentiating between the effector and complex, and may be used to determine enzymatic function such as by identifying or quantifying the substrate and product.

A screening assay in which candidate agent binding of cofactors is analysed can include a number of conditions. These conditions include but are not limited to pH, temperature, tonicity, the presence of relevant other proteins, and relevant modifications to the polypeptide such as glycosylation or lipidation. It is contemplated that the cofactors can be expressed and utilized in a prokaryotic or eukaryotic cell. The host cell expressing the cofactors can be used whole or the cofactor can be isolated from the host cell. The cofactor can be membrane bound in the membrane of the host cell or it can be free in the cytosol of the host cell. The host cell can also be fractionated into sub-cellular fractions where the cofactor can be found. For example, cells expressing the cofactor can be fractionated into the nuclei, the *endoplasmic reticulum*, vesicles, or the membrane surfaces of the cell.

pH is preferably from about a value of 6.0 to a value of about 8.0, more preferably from about a value of about 6.8 to a value of about 7.8, and most preferably, about 7.4. In a preferred embodiment, temperature is from about 20°C degrees to about 50°C degrees more preferably, from about 30°C degrees to about 40°C degrees and even more preferably about 37°C de-

grees. Osmolality is preferably from about 5 milliosmols per liter (mosm/L) to about 400 mosm/l, and more preferably, from about 200 milliosmols per liter to about 400 mosm/l and, even more preferably from about 290 mosm/L to about 310 mosm/L. The presence of further cofactors or other proteins can be required for the proper functioning of the cofactors according to the invention. Typical chemical cofactors include sodium, potassium, calcium, magnesium, and chloride. In addition, small, non-peptide molecules, known as prosthetic groups may also be required. Other biological conditions needed for cofactor function are well-known in the art.

It is well-known in the art that proteins can be reconstituted in artificial membranes, vesicles or liposomes. (Danboldt et al.,1990). The present invention contemplates that the cofactor can be incorporated into artificial membranes, vesicles or liposomes. The reconstituted cofactor can be utilized in screening assays.

It is further contemplated that a cofactor of the present invention can be coupled to a solid support, e.g., to agarose beads, polyacrylamide beads, polyacrylic, sepharose beads or other solid matrices capable of being coupled to polypeptides. Well-known coupling agents include cyanogen bromide (CNBr), carbonyldiimidazole, tosyl chloride, diaminopimelimidate, and glutaraldehyde.

In a typical screening assay for identifying candidate substances, one employs the same recombinant expression host as the starting source for obtaining the cofactor polypeptide, generally prepared in the form of a crude homogenate. Recombinant cells expressing the cofactor are washed and homogenized to prepare a crude polypeptide homogenate in a desirable buffer such as disclosed herein. In a typical assay, an amount of polypeptide from the cell homogenate, is placed into a small volume of an appropriate assay buffer at an appropriate pH. Candidate substances, such as agonists and antagonists, are added to the admixture in convenient concentrations and the interaction between the candidate substance and the cofactor polypeptide is monitored.

Where one uses an appropriate known substrate for the cofactors, one can, in the foregoing manner, obtain a baseline activity for the recombinantly produced cofactors. Then, to test for inhibitors or modifiers of the cofactor function, one can incorporate into the admixture a candidate substance whose effect on the cofactor is unknown. By comparing reactions which are

carried out in the presence or absence of the candidate substance, one can then obtain information regarding the effect of the candidate substance on the normal function of the cofactor.

Accordingly, this aspect of the present invention will provide those of skill in the art with methodology that allows for the identification of candidate substances having the ability to modify the action of cofactor polypeptides in one or more manners.

Additionally, screening assays for the testing of candidate substances are designed to allow the determination of structure-activity relationships of agonists or antagonists with the cofactors, e.g., comparisons of binding between naturally-occurring hormones or other substances capable of interacting with or otherwise modulating the cofactor; or comparison of the activity caused by the binding of such molecules to the cofactor.

In certain aspects, the polypeptides of the invention are crystallized in order to carry out x-ray crystallographic studies as a means of evaluating interactions with candidate substances or other molecules with the cofactor polypeptide. For instance, the purified recombinant polypeptides of the invention, i.e. of the cofactors according to the invention, when crystallized in a suitable form, are amenable to detection of intra-molecular interactions by x-ray crystallography. In another aspect, the structure of the polypeptides can be determined using nuclear magnetic resonance.

PHARMACEUTICAL COMPOSITION:

This invention provides a pharmaceutical composition comprising an effective amount of an agonist or antagonist drug identified by the method described herein and a pharmaceutically acceptable carrier. Such drugs and carrier can be administered by various routes, for example oral, subcutaneous, intramuscular, intravenous or intracerebral. The preferred route of administration would be oral at daily doses of about 0.01 -100 mg/kg.

This invention provides a method of treating diseases such as cancer, cardiovascular diseases, bone diseases, hormonal dysfunctions and others by altering the activity of the cofactor thereby influencing the binding affinity of the cofactor to ER alpha.

TRANSFORMATION OF CELLS AND DRUG SCREENING :

The recombinant expression constructs of the present invention are useful in molecular biology to transform cells which do not ordinarily express the CF16 to CF19 and CF40 to CF43 polypeptides to express these cofactors upon transformation.

Such cells are useful as intermediates for making cellular preparations useful for cofactor binding assays, which are in turn useful for drug screening.

The recombinant expression constructs of the present invention are also useful in gene therapy. Cloned genes of the present invention, or fragments thereof, may also be used in gene therapy carried out by homologous recombination or site-directed mutagenesis. See generally Thomas & Capecchi, Cell 51, 503-512 (1987); Bertling, Bioscience Reports 7, 107-112 (1987); Smithies et al., Nature 317, 230-234 (1985).

Oligonucleotides of the present invention are useful as diagnostic tools for probing cofactor gene expression in tissues. For example, tissues are probed *in situ* with oligonucleotide probes carrying detectable groups by conventional autoradiographic techniques, as explained in greater detail in the examples below, to investigate native expression of this cofactor or pathological conditions relating thereto. Further, chromosomes can be probed to investigate the presence or absence of the CF genes, and potential pathological conditions related thereto, as also illustrated by the Examples below. Probes according to the invention should generally be at least about 15 nucleotides in length to prevent binding to random sequences, but, under the appropriate circumstances may be smaller.

ANTIBODIES AGAINST THE CF16, CF17, CF18, CF19, CF40, CF41, CF42, and/or CF43 COFACTOR PROTEIN OR POLYPEPTIDE

Another aspect of the invention includes antibodies specifically reactive with the proteins or any parts of the proteins according to the invention and or polypeptides encoded by the nucleotide sequences of the cofactors. The term „antibody" refers to intact molecules as well as fragments thereof, such as Fa, F(ab).sub.2, and Fv, which are capable of binding the epitopic determinant. By using immunogens derived from the polypeptide according to the invention and/or encoded by the nucleic acids according to the invention, anti-protein/anti-peptide an-

tiserum or monoclonal antibodies can be made by standard protocols (E. Howell & D. Lane. Antibodies: A Laboratory Manual. *Cold Spring Harbor Laboratory* (1988)).

A polyclonal antibody is prepared by immunizing a mammal, such as a mouse, a hamster or rabbit with an immunogenic form of the cofactors according to the invention depending on which of these are desired) of the present invention, and collecting antisera from that immunized animal. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As an immunizing antigen, fusion proteins, intact polypeptides or fragments containing small peptides of interest can be used. They can be derived by expression from a cDNA transfected in a host cell with subsequent recovering of the protein/peptide or peptides can be synthesized chemically (*e.g.* oligopeptides with 10-15 residues in length). Important tools for monitoring the function of the cofactor gene according to the present invention, *i.e.* encoded by a sequence according to SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, or SEQ ID NO. 22 are antibodies against various domains of the proteins according to the invention.

A given polypeptide or polynucleotide may vary in its immunogenicity. It is often necessary to couple the immunogen (e.g. the polypeptide) with a carrier. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin (BSA) and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal in the presence of an adjuvant, a non-specific stimulator of the immune response in order to enhance immunogenicity. The production of polyclonal antibodies is monitored by detection of antibody titers in plasma or serum at various time points following immunization. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. When a desired level of immunogenicity is obtained, the immunized animal may be bled and the serum isolated, stored and purified.

To produce monoclonal antibodies, antibody-producing cells (e.g. spleen cells) from an immunized animal (preferably mouse or rat) are fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Where the immunized animal is a mouse, a preferred myeloma cell is the murine NS-1 myeloma cell. Such techniques are well known in the art, and include, for example, the hybridoma technique

(originally developed by Kohler & Milstein. *Nature* 256: 495-497 (1975)), the human B cell hybridoma technique (Kozbar *et al. Immunology Today* 4:72 (1983)), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.* Monoclonal Antibodies and Cancer Therapy. *Alan R. Liss, Inc.* pp. 77-96 (1985)).

The fused spleen/myeloma cells are cultured in a selective medium to select fused spleen/myeloma cells from the parental cells. Fused cells are separated from the mixture of non-fused parental cells, for example, by the addition of agents that block the *de novo* synthesis of nucleotides in the tissue culture media. This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants for reactivity with an antigen-polypeptide. The selected clones may then be propagated indefinitely to provide the monoclonal antibody in convenient quantity.

The creation of antibodies which specifically bind the polypeptides according to the invention and/or encoded by the nucleotide sequences of the cofactors or their complements provides an important utility in immunolocalization studies, and may play an important role in the diagnosis and treatment of such diseases and disorders as metabolic disorders, immunological indications, hormonal dysfunctions and/or neurosystemic diseases. The antibodies may be employed to identify tissues, organs, and cells which express the cofactors. Antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate cofactor protein levels in tissue or from cells in bodily fluid as part of a clinical testing procedure.

Monoclonal antibodies provided by the present invention are also produced by recombinant genetic methods well known to those of skill in the art, and the present invention encompasses antibodies made by such methods that are immunologically reactive with an epitope of a mammalian cofactor protein or peptide according to the invention.

The present invention encompasses fragments of the antibody that are immunologically reactive with an epitope of a cofactor protein or peptide. Such fragments are produced by any number of methods, including but not limited to proteolytic cleavage, chemical synthesis or preparation of such fragments by means of genetic engineering technology. The present in-

vention also encompasses single-chain antibodies that are immunologically reactive with an epitope of a cofactor protein or peptide made by methods known to those of skill in the art.

CHIMERIC ANTIBODIES AND OTHER TYPES OF ANTIBODIES:

The invention also includes chimeric antibodies, comprised of light chain and heavy chain peptides immunologically reactive to an epitope that is a cofactor protein or peptide according to the invention. The chimeric antibodies embodied in the present invention include those that are derived from naturally occurring antibodies as well as chimeric antibodies made by means of genetic engineering technology well known to those of skill in the art as, for example, humanized antibodies. Furthermore, the antibodies can also be chemically and/or enzymatically modified, for example carry a glycosylation and/or a label, like a fluorescent or radioactive label.

Also included are methods for the generation of antibodies against any of the group comprising the peptides according to SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12, SEQ ID NO. 15, SEQ ID NO. 18, SEQ ID NO. 21, and/or SEQ ID NO. 24 which rely on the use of phage display systems and related systems, such as described in Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC, Immunotechnology 1998 Jun;4(1):1-20, and references therein.

EPITOPES OF THE CF16, CF17, CF18, CF19, CF40, CF41, CF42, and CF43 COFACTORS

The present invention also encompasses one or more epitopes of a cofactor protein or peptide that is comprised of sequences and/or a conformation of sequences present in the cofactor proteins or peptide molecule. These epitopes may be naturally occurring, or may be the result of proteolytic cleavage of the cofactor proteins or peptides and isolation of an epitope-containing peptide or may be obtained by synthesis of an epitope-containing peptide using a method of genetic engineering technology and synthesized by genetically engineered prokaryotic or eukaryotic cells.

ANTISENSE OLIGONUCLEOTIDES AGAINST CF16, CF17, CF18, CF19, CF40, CF41, CF42 and CF43 GENE TRANSCRIPTS

Antisense oligonucleotides are short single stranded DNA or RNA molecules which may be used to block the availability of the cofactor messenger(s). Synthetic derivatives of ribonucleotides or desoxyribonucleotides and/or PNAs (see above) are equally possible. These are potential candidate agents which may interact with the cofactor according to the invention.

The sequence of an antisense oligonucleotide is at least partially complementary to the sequence of the cofactor of interest. The complementarity of the sequence is in any case high enough to enable the antisense oligonucleotide to bind to the nucleic acid according to the invention or parts thereof (SEQ ID NO1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 13, SEQ ID NO. 16, SEQ ID NO. 19, and/or SEQ ID 22) in which the binding of oligonucleotides to the target sequence interfere with the biological function of the targeted sequence (Brysch W, Schlingensiepen KH, Design and application of antisense oligonucleotides in cell culture, in vivo, and as therapeutic agents, Cell Mol Neurobiol 1994 Oct;14(5):557-68; Wagner RW, Gene inhibition using antisense oligodeoxynucleotides, Nature 1994 Nov 24;372(6504):333-5 or Brysch W, Magal E, Louis JC, Kunst M, Klinger I, Schlingensiepen R, Schlingensiepen KH Inhibition of p185c-erbB-2 proto-oncogene expression by antisense oligodeoxynucleotides down-regulates p185-associated tyrosine-kinase activity and strongly inhibits mammary tumor-cell proliferation, Cancer Gene Ther 1994 Jun;1(2):99-105 or Monia BP, Johnston JF, Ecker DJ, Zounes MA, Lima WF, Freier SM Selective inhibition of mutant Ha-ras mRNA expression by antisense oligonucleotides, J Biol Chem 1992 Oct 5;267(28):19954-62 or Bertram J, Palfner K, Killian M, Brysch W, Schlingensiepen KH, Hiddemann W, Kneba M, Reversal of multiple drug resistance in vitro by phosphorothioate oligonucleotides and ribozymes, Anticancer Drugs 1995 Feb;6(1):124-34)

This interference occurs in most instances at the level of translation, *i.e.* through the inhibition of the translational machinery by oligonucleotides that bind to mRNA, however, two other mechanisms of interference with a given gene's function by oligonucleotides can also be envisioned, (i) the functional interference with the transcription of a gene through formation of a triple helix at the level of genomic DNA and the interference of oligonucleotides with the function of RNA molecules that are executing at least part of their biological function in the untranslated form (Kochetkova M, Shannon MF, Triplex-forming oligonucleotides and their use in the analysis of gene transcription. Methods Mol Biol 2000;130:189-201 Rainer B.

Lanz1, Neil J. McKenna1, Sergio A. Onate1, Urs Albrecht2, Jiemin Wong1, Sophia Y. Tsai1, Ming-Jer Tsai1, and Bert W. O'Malley A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex Cell, Vol. 97, 17–27, April, 1999).

Antisense oligonucleotides can be conjugated to different other molecules in order to deliver them to the cell or tissue expressing any of the cofactor genes. For instance the antisense oligonucleotide can be conjugated to a carrier protein (*e.g.* ferritin) in order to direct the oligonucleotide towards the desired target tissue, *i.e.* in case of ferritin predominantly to the liver.

Antisense expression constructs are expression vector systems that allow the expression – either inducible or uninducible - of a complementary sequence to the CF16 to CF19 and CF40 to CF43 cofactor sequences according to the invention. The potential possibility of such an approach has been demonstrated in many different model systems (von Ruden T, Gilboa E, Inhibition of human T-cell leukemia virus type I replication in primary human T cells that express antisense RNA, J Virol 1989 Feb;63(2):677-82; Nemir M, Bhattacharyya D, Li X, Singh K, Mukherjee AB, Mukherjee BB, Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency, J Biol Chem 2000 Jan 14;275(2):969-76; Ma L, Gauville C, Berthois Y, Millot G, Johnson GR, Calvo F Antisense expression for amphiregulin suppresses tumorigenicity of a transformed human breast epithelial cell line, Oncogene 1999 Nov 11;18(47):6513-20; Refolo LM, Eckman C, Prada CM, Yager D, Sambamurti K, Mehta N, Hardy J, Younkin SG, Antisense-induced reduction of presenilin 1 expression selectively increases the production of amyloid beta42 in transfected cells, J Neurochem 1999 Dec;73(6):2383-8; Buckley NJ, Abogadie FC, Brown DA, Dayrell M, Caulfield MP, Delmas P, Haley JE, Use of antisense expression plasmids to attenuate G-protein expression in primary neurons, Methods Enzymol 2000;314:136-48).

According to the invention an antisense expression construct can be constructed with virtually any expression vector capable of fulfilling at least the basic requirements known to those skilled in the art.

In one embodiment of the invention retroviral expression systems or tissue specific gene expression systems are preferred.

Current standard technologies for delivering antisense constructs are performed through a conjugation of constructs with liposomes and related, complex-forming compounds, which are transferred via electroporation techniques or via particle-mediated "gene gun" technologies into the cell. Other techniques may be envisioned by one skilled in the art.

Microinjection still plays a major role in most gene transfer techniques for the generation of germ-line mutants expressing foreign DNA (including antisense RNA constructs) and is preferred embodiment of the present invention.

RIBOZYMES DIRECTED COFACTOR 16 TO 19 and CF40 TO CF43 GENE TRANSCRIPT.

Ribozymes are either RNA molecules (Gibson SA, Pellenz C, Hutchison RE, Davey FR, Shillitoe EJ, Induction of apoptosis in oral cancer cells by an anti-bcl-2 ribozyme delivered by an adenovirus vector, Clin Cancer Res 2000 Jan;6(1):213-22; Folini M, Colella G, Villa R, Lualdi S, Daidone MG, Zaffaroni N, Inhibition of Telomerase Activity by a Hammerhead Ribozyme Targeting the RNA Component of Telomerase in Human Melanoma Cells, J Invest Dermatol 2000 Feb;114(2):259-267; Halatsch ME, Schmidt U, Botefur IC, Holland JF, Ohnuma T, Marked inhibition of glioblastoma target cell tumorigenicity in vitro by retrovirus-mediated transfer of a hairpin ribozyme against deletion-mutant epidermal growth factor receptor messenger RNA, J Neurosurg 2000 Feb;92(2):297-305; Ohmichi T, Kool ET, The virtues of self-binding: high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes, Nucleic Acids Res 2000 Feb 1;28(3):776-783) or DNA molecules (Li J, Zheng W, Kwon AH, Lu Y, In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme; Nucleic Acids Res 2000 Jan 15;28(2):481-488) that have catalytic activity. The catalytic activity located in one part of the RNA (or DNA) molecule can be "targeted" to a specific sequence of interest by fusing the enzymatically active RNA molecule sequence with a short stretch of RNA (or DNA) sequence that is complementary to the cofactor gene transcript of interest. Such a construct will, when introduced into a cell either physically or via gene transfer of a ribozyme expression construct find the corresponding cofactor sequence (our sequence of interest or also targeted in RNA) and bind via its sequence-specific part to said sequence. The catalytic activity attached to the construct, usually associated with a special nucleic acid structure (people distinguish so called "hammerhead" structures and "hairpin" structures), will then cleave the targeted RNA. The targeted mRNA will be destroyed and cannot be translated efficiently, thus the protein encoded by the

mRNA derived from cofactor will not be expressed or at least will be expressed at significantly reduced amounts.

These are potential candidate agents which may interact with the cofactor according to the invention.

In a preferred embodiment the invention covers inducible ribozyme constructs (Koizumi M, Soukup GA, Kerr JN, Breaker RR, Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP, Nat Struct Biol 1999 Nov;6(11):1062-1071).

In a further preferred embodiment the invention concerns the use of "bivalent" ribozymes (multimers of catalytically active nucleic acids) as described in (Hamada M, Kuwabara T, Warashina M, Nakayama A, Taira K, Specificity of novel allosterically trans- and cis-activated connected maxizymes that are designed to suppress BCR-ABL expression FEBS Lett 1999 Nov 12;461(1-2):77-85).

TRANSGENIC ANIMALS CARRYING THE CF16, CF17, CF18, CF19, CF40, CF41, CF42 AND/OR CF43 GENE

Also provided by the present invention are non-human transgenic animals grown from germ cells transformed with a CF16, CF17, CF18, CF19, CF40, CF41, CF42 OR CF43 nucleic acid sequence according to the invention and that express the cofactor according to the invention and offspring and descendants thereof. Also provided are transgenic non-human mammals comprising a homologous recombination knockout of the native cofactors, as well as transgenic non-human mammals grown from germ cells transformed with nucleic acid antisense to the nucleic acids of the invention and offspring and descendants thereof. Further included as part of the present invention are non-human transgenic animals in which the native cofactor has been replaced with the human ortholog. Of course, offspring and descendants of all of the foregoing transgenic animals are also encompassed by the invention.

Transgenic animals according to the invention can be made using well known techniques with the nucleic acids disclosed herein. E.g., Leder et al., U.S. Patent Nos.4,736,866 and 5,175,383; Hogan et al., Manipulating the Mouse Embryo, A Laboratory Manual (Cold Spring Harbor Laboratory (1986)); Capecchi, Science 244, 1288 (1989); Zimmer and Gruss, Nature 338, 150 (1989); Kuhn et al., Science 269, 1427 (1995); Katsuki et al., Science 241,

593 (1988); Hasty et al., Nature 350, 243 (1991); Stacey et al., Mol. Cell Biol. 14, 1009 (1994); Hanks et al., Science 269, 679 (1995); and Marx, Science 269, 636 (1995). Such transgenic animals are useful for screening for and determining the physiological effects of the cofactor agonists and antagonist.

Consequently, such transgenic animals are useful for developing drugs to regulate physiological activities in which the cofactors participate.

MODELLING OF THE STRUCTURE OF CF16, CF17, CF18, CF19, CF40, CF41, CF42 AND/OR CF43.

In one embodiment of the invention the amino acid sequences of the present invention can be used for structural drug design. Aim is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g. agonists, antagonists or inhibitors) in order to design drugs which are, for example, more active or stable forms of the polypeptide, or which, for example, enhance or interfere with the function of a polypeptide *in vivo*. In one approach one first determines the three-dimensional structure of a protein of interest, *i.e.* the cofactor, by computer-modeling, x-ray crystallography or a combination of both approaches. Additional useful information with respect to the structure of a polypeptide could also be gained from comparison of the protein sequence of the protein of interest with the sequence of related proteins where the structure is known. From the three-dimensional structure, binding sites of potential inhibitors or activators can be predicted. It can further be predicted which kinds of molecule might bind there. The predicted substances can then be screened to test their effect on the activity of the protein and its biological function.

The invention is further illustrated by the following figure and examples from which further features, advantages and embodiments can be taken. The following Examples are provided for illustrative purposes only and are not intended, nor should they be construed, as limiting the invention in any manner.

Figure 1 shows yeast-two-hybrid interactions of CF16 tested against a set of nuclear receptors and cofactors (see Example 4). A Gal4-DNA binding domain-CF16 fusion protein was tested for interactions against a panel of Gal4 activation domain fusion proteins as described in the text. Values for interactions are shown as fold activation over the CF16 interaction with the

Gal4 activation domain only (protein 1-1 upper panel and protein 13-1 lower panel). Gal4 activation domain fusion proteins tested are:

(LBD-ligand binding domain; frag.-fragment; NT-N-terminus; CT-C-terminus

1-1	Gal4 activation domain		2-1	RORalpha-LBD		3-1	ERbeta
1-2	TRalpha-LBD	2-2	RXRalpha-LBD		3-2	ERRg-LBD	
1-3	AR-LBD	2-3	RARalpha	3-3	VDR-LBD		
1-4	PPARgamma-LBD	2-4	RXRalpha	3-4	GRalpha-LBD		
1-5	PPARgamma	2-5	SHP	3-5	GRalpha		
1-6	PPARalpha	2-6	ERalpha-LBD	3-6	GRalpha-NT		
1-7	PXR-LBD	2-7	ERalpha		3-7	LXRalpha	
1-8	PR	2-8	ERbeta-LBD	3-8	LXRalpha-LBD		
4-1	LXRbeta	5-1	SRCI	6-1	TRIP1		
4-2	LXRbeta-LBD		5-2	JAB1		6-2	COX1
4-3	FXRalpha-LBD		5-3	TIF2		6-3	TIP 60
4-4	Lion1	5-4	CBP frag.	6-4	ALIEN		
4-5	NcoA3	5-5	NCOA62	6-5	FLH2		
4-6	TRAP220 frag.		5-6	PCAF		6-6	ARA 55
4-7	TRAP220	5-7	RAP250		6-7	ARA 70	
4-8	SRCI frag.	5-8	DRIP150	6-8	TAFII250		
7-1	SunCOR	8-1	CalNUC		9-1	POB1	
7-2	Lion2	8-2	Lion3	9-2	L7 SPA		
7-3	NCOR1		8-3	CF16		9-3	NURR1 LBD
7-4	SMRT	8-4	Lion4	9-4	HNF4 g-LBD		
7-5	RIP140		8-5	HNF4A-LBD	9-5	ARA55-CT	
7-6	TSG101	8-6	PGCI	9-6	PPARbeta-LBD		
7-7	NEFA	8-7	MIP224		9-7	TR4-0R-LBD	
7-8	NEFA NT	8-8	PHLP	9-8	EAR1		
10-1	Lion5	11-1	COUP-TFII -LBD		12-1	PPARbeta	
10-2	NRIF3	11-2	CAR1-LBD	12-2	HREQ		
10-3	RXRg	11-3	NOR1-LBD	12-3	CAR1		
10-4	PGCI-CT	11-4	REA	12-4	ERR1		
10-5	EAR1-LBD	11-5	LRH1-LBD	12-5	TLX-LBD		
10-6	Lion6	11-6	MR-LBD	12-6	empty		
10-7	Lion7	11-7	TR2-11-LBD	12-7	COUP-TFII		
10-8	ERRalpha-LBD		11-8	RORalpha-LBDtrunc	12-8	SF1	
13-1	Gal4 activation domain		14-1	MIF1 frag.	15-1	Lion8	
13-2	ERR3	14-2	CIB frag		15-2	Lion9	
13-3	RORbeta-LBD		14-3	KIAA1106	15-3	Lion10	
13-4	SHP	14-4	KIAA0535	15-4	CIB		
13-5	CIA	14-5	PGC1-CT	15-5	Lion11		
13-6	Lion12	14-6	PGC1-NT	15-6	HNF4g		
13-7	EC2.3.16	14-7	KIDlike	15-7	Lion13		
13-8	EC2.3.16 frag	14-8	DAX1trunc	15-8	Lion14		
16-1	Lion15	16-2	SF1-LBD	16-3	RORalpha		
16-4	ERRg-LBD	16-5	DAX1				

SEQUENCE DESCRIPTIONS:

SEQ ID No 1 to **3** show **CF16** (tremblnewIAK023173) sequences, in particular the cDNA Sequence (Seq ID No 1) the Reverse Complement of the cDNA Sequence (Seq ID No 2) and the amino acid sequence (Seq ID No 3),

SEQ ID No 4 to **6** show **CF17** (nageneseqIZ41321) sequences, in particular the cDNA Sequence (Seq ID No 4) the Reverse Complement of the cDNA Sequence (Seq ID No 5) and the amino acid sequence (Seq ID No 6),

SEQ ID No 7 to **9** show **CF18** (Nkx2.2(aageneseqiY25173)) sequences, in particular the cDNA Sequence (Seq ID No 7) the Reverse Complement of the cDNA Sequence (Seq ID No 8) and the amino acid sequence (Seq ID No 9),

SEQ ID No 10 to **12** show **CF19** (CAM2(tremblIAF112472)) sequences, in particular the cDNA Sequence (Seq ID No 10) the Reverse Complement of the cDNA Sequence (Seq ID No 11) and the amino acid sequence (Seq ID No 12),

SEQ ID No 13 to **15** show **CF40** sequences (NR26BP5, NM_014958), in particular the cDNA Sequence (Seq ID No 13) the Reverse Complement of the cDNA Sequence (Seq ID No 14) and the amino acid sequence (Seq ID No 15),

SEQ ID No 16 to **18** show **CF41** sequences (NR26BP7; aageneseq|Y94906|Y94906), in particular the cDNA Sequence (Seq ID No 16) the Reverse Complement of the cDNA Sequence (Seq ID No 17) and the amino acid sequence (Seq ID No 18),

SEQ ID No 19 to **21** show **CF42** sequences (NR51BP1, nageneseq|C76971|C76971 Human ORFX ORF2526 polynucleotide), in particular the cDNA Sequence (Seq ID No 19) the Reverse Complement of the cDNA Sequence (Seq ID No 20) and the amino acid sequence (Seq ID No 21),

SEQ ID No 22 to **24** show **CF43** sequences (NR51BP2, embl|T11452|HST11452 CHR90018 Chromosome 9 exon), in particular the cDNA Sequence (Seq ID No 22) the Reverse Complement of the cDNA Sequence (Seq ID No 23) and the amino acid sequence (Seq ID No 24),

SEQ ID No 25 to **27** show ER alpha sequences, in particular the cDNA Sequence (Seq ID No 25) the Reverse Complement of the cDNA Sequence (Seq ID No 26) and the amino acid sequence of the full length ER alphy bait (Seq ID No 27),

SEQ ID No 28 to **30** show ER alpha bait sequences, in particular the cDNA Sequence (Seq ID No 28) the Reverse Complement of the cDNA Sequence (Seq ID No 29) and the amino acid sequence of the ER alpha ligand binding domain bait (Seq ID No 30),

EXAMPLES

EXAMPLE 1: CLONING AND EXPRESSION OF THE GENES ACCORDING TO THE INVENTION

Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques that are well understood in the art. Isolated plasmids, DNA sequences, or synthesised oligonucleotides are cleaved, tailored, and religated in the form desired.

Site-specific DNA cleavage is performed by treatment with the suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes. See, e.g., New England Biolabs, Product Catalog. In general, about 1 µg of plasmid and/or DNA sequence is cleaved by one unit of enzyme in about 20 µl of buffer solution. Often excess of restriction enzyme is used to ensure complete digestion of the DNA substrate. Incubation times of about one hour to two hours at about 37°C are workable, although variations are tolerable.

After each incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether extraction. The nucleic acid may be recovered from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in Methods in Enzymology 65, 499-560 (1980).

Transformed host cells are cells which have been transformed or transfected with recombinant expression constructs made using recombinant DNA techniques and comprising cofactor encoding sequences. Preferred host cells for transient transfection are COS-7 cells. Transformed host cells may ordinarily express one of the CF16 to CF19 and CF40 to CF43 cofactors but host cells transformed for purposes of cloning or amplifying nucleic acid hybridisation probe DNA need not express the cofactors. When expressed, the cofactor proteins will typically be located in the host cell membrane.

Cultures of cells derived from multicellular organisms are desirable hosts for recombinant nuclear receptor protein synthesis. In principal, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. However, mammalian cells are preferred. Propagation of such cells in cell culture has become a routine procedure. See Tissue Culture (Academic Press, Kruse & Patterson, Eds., 1973). Examples of useful host cell lines are bacteria cells, insect cells, yeast cells, human 293 cells, VERO and HeLa cells, LMTK- cells, and WI138, BHK, COS-7, CV, and MDCK cell lines. Human 293 cells are preferred.

EXAMPLE 2: COFACTOR TISSUE LOCALIZATION:

A multiple tissue northern blot (Clontech, Palo Alto) is hybridized to a labeled probe. The blot contains about 0.3 to 3 μg of poly A RNA derived from various tissues. Hybridization may be carried out in a hybridization solution such as one containing SSC (see Maniatis et al, ibid) at an optimized temperature between 50°c and 70°C, preferably 65°C. The filter may be washed and a film exposed for signal detection (see also: Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

EXAMPLE 3: COFACTOR cDNA ISOLATION FROM HUMAN AND OTHER ORGANISMS:

A cloning strategy is used to clone the desired CF cofactor cDNA from specific cDNA libraries (Clontech, Palo Alto) or alternatively, RNA is obtained from various tissues and used to prepare cDNA expression libraries by using for example an Invitrogen kit. (Invitrogen Corporation, San Diego). For the isolation of the CF cDNA clones the chosen library may be screened under stringent condition (see definitions above) by using CF16, CF17, CF18, CF19, CF40, CF41, CF42 or CF43 specific probes. The cDNA insert of positive clones is subsequently sequenced and cloned in a suitable expression vector.

Additionally, full length cofactor clones from various species are obtained by using RACE PCR technology. In brief, suitable cDNA libraries are constructed or purchased. Following reverse transcription, the first strand cDNA is used directly in RACE PCR reactions using a RACE cDNA amplification kit according to the manufactures protocol (Clontech, Palo Alto). Amplified fragments are purified, cloned and subsequently used for sequence analysis.

In order to obtain information about the genomic organization of the cofactor gene, genomic libraries (Clontech, Palo Alto) are screened with a receptor specific probe under stringent conditions. Positive clones are isolated and the complete DNA sequence of the putative receptor is determined by sequence analysis (Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

EXAMPLE 4: ISOLATION OF THE COFACTOR PROTEINS BY USE OF THE YEAST TWO-HYBRID SYSTEM

A yeast two-hybrid assay was performed using methods such as described by Fields and Song *Nature* 340, pp245 (1989), Bartel et al., *Biotechniques* 14, pp920 (1993) and Lee et al. Nature 374 pp91-4 (1995). A sequence encoding amino acids (aa) 249-595 of ER alpha (containing the ligand binding domain; LBD) or alternatively a sequence encoding the full length ER alpha protein (amino acids 1-595) was cloned into the vector pGBT9 (Clontech) in such way that, after transformation of the haploid yeast strain CG1945 (Clontech), a hybrid protein is expressed consisting of the DNA-binding domain (BD) of the Gal4 transcription factor fused N-terminally to amino acids 249-595 of ER alpha (SEQ ID NO. 33) or to amino acids 1-595 of ER alpha (SEQ ID NO. 30), respectively. CG1945 cells expressing the Gal4BD::ER alpha(aa 249-595) or the Gal4BD::ER alpha (aa1-595) fusion protein were mated to cells of strain Y187 (Clontech) containing a library of Gal4 transcription activation domain (AD) fusion plasmids with human cDNA generated from a range of tissues inserted into the vector pACT2 (Clontech). All libraries were purchased from Clontech Laboratories (MATCHMAKER human cDNA libraries) and included Cat. numbers HL4040AH (aorta), HL4041AH (chondrocytes), HY4004AH (brain), HY4035AH (testis), HY4024AH (liver), HY4042AH (heart), HY4053AH (bone marrow), HY4028AH (fetal brain), HY4043AH (kidney), HY4051AH (ovary), HY4047AH (skeletal muscle) and HY4000AA (hela). The two-hybrid screens were essentially performed following the Clontech "Pretransformed Matchmaker Libraries User Manual" (PT3183-1): Transformed CG1945 and Y187 cells were mated

in order to coexpress the Gal4BD::ER alpha (aa249-595) fusion protein or the Gal4BD::ER alpha(aa1-595) fusion protein, respectively, and the Gal4AD fusion proteins encoded on the library plasmids within one cell.

Interaction of the two hybrid proteins led to activation of reporter gene transcription. Cells were selected for interactions of ER alpha with library proteins in medium lacking tryptophan, leucine and histidine and containing 40 mM 3-aminotriazol as well as 400 nM 17ß-estradiol and were further assayed for expression of □-galactosidase, encoded by the *MEL1* reporter gene. Colonies which were positive for reporter gene activation were chosen for further analysis. The DNA inserts of the library plasmids contained in these colonies were amplified by use of the polymerase chain reaction directly on the yeast colonies using oligonucleotide primers which hybridize on vector sequences flanking both sides of the insert. The identity of the insert was determined by standard DNA sequencing techniques.

Six novel cofactors (CF16, CF17, CF18, CF19, CF40, and CF41) interacting with the Gal4BD::ER alpha(aa249-595) fusion protein were isolated using this approach: CF16 was isolated from the aorta, bone marrow, testis, skeletal muscle and brain cDNA libraries, CF17 from the brain and testis libraries, CF18 and CF19 from the brain cDNA library, CF40 (NR26BP5) from the liver and kidney cDNA libraries and CF41 (NR26BP7) from the kidney cDNA library.

When using the Gal4BD::ER alpha(aa1-595) fusion protein two novel cofactors were isolated: CF42 (NR51BP1) was isolated from the heart and aorta libraries and CF43 (NR51BP2) was isolated from the brain and aorta libraries.

Alternatively, the yeast two hybrid approach could be set up in a more directed manner, in such way that a DNA fragment encoding the full length CF16 protein is cloned into the vector pGBT9 so that a fusion protein is expressed in which CF16 is fused with its N-terminus to the Gal4 DNA binding domain. The vector is then transformed into yeast strain CG1945 (Clontech). The cells are grown at 30 ^{O}C to an optical density (OD 600 nm) of 1.0. In parallel a number of protein encoding fragments or full length open reading frames (see legend Fig.XXX) are cloned into the vector pGAD424 (Clontech) in such way that proteins are expressed which are fused at their N-termini to the Gal4 transcription activation domain. The resulting plasmids are transformed into yeast strain Y187 (Clontech). The Y187-

transformants are grown as described above. After that 25 µl aliquots of the cultures become mixed in the wells of a 96 well microtiterplate with 25 µl aliquots of the CG1945 transformants containing the CF16 DNA binding domain fusion protein. 50 µl of selective medium lacking leucine and tryptophan (SD-LW) and containing 10 % YPDA rich medium is added to each well to improve mating efficiency. Cells are left overnight at 30 ^{o}C for mating. The next day 5 µl is transferred from the wells into a fresh microtiter plate containing 150 µl of selective -LW medium per well. The plate is incubated for two days at 30 ^{o}C in order to let the resulting diploids grow to saturation. The latter transfer is repeated once and again cells are incubated for two days as above. Eventually 3 µl are transferred from each well into 100 µl of selective medium lacking leucine, tryptophan and histidine (SD-LWH) containing 50µM of 4-Methylumbelliferyl-alpha-D galactoside (4-Mu-X). To test for a ligand dependence of the interactions alternatively the selective medium contains in addition 17 ß-estradiol in 250 nM concentration. The cells are incubated for exactly 72 hours at 30 ^{o}C in order to select for histidine prototroph clones i.e. expression of Gal4 activation domain fusion proteins interacting with CF16. In addition, *MEL1* reporter gene activation is tested via fluorimetric detection of 4-Methylumbelliferone (4-Mu) in the medium. 4-Mu is one of the products resulting from cleavage of 4-Mu-X by alpha-galactosidase, the *MEL1* gene product. 4-Mu emits fluorescent light of 465 nm wavelength, if excited with light of 360 nm wavelength. Thus the fluorescence units measured in each well give an indication on the relative strength of the interaction of a given Gal4 activation domain fusion protein with CF16. It turns out that amongst the proteins tested, the ligand binding domain of ERalpha interacted in a strictly estradiol dependent fashion with CF16 (Fig.1; compare interactions 2-6 in A and 2-6 in B). A strong interaction with the ligand binding domain of the estrogen related receptor alpha (ERR1) was also observed (10-8 A) and this interaction could be stimulated weakly by estradiol (10-8 B). Furthermore, CF16 interacted significantly with ERR3, but this interaction is not influenced by estradiol (13-2 A and B). The ligand binding domain of the liver receptor homologue (LRH-1) was also found to interact weakly but estradiol independently with CF16 (11-5 A and B).

The experiment was repeated but instead of estradiol a set of other chemical molecules known to influence the function of specific nuclear receptors was added to the medium: these included rifampicin (in 1 µM concentration), vitamin D3 (1 µM), all-trans retinoic acid (1 µM), 9-cis retinoic acid (1 µM), dexamethasone (0.5 µM), androstane (0.5 µM), linoleic acid (10 µM), aldosterone (0.5 µM), triiodothyronine (1 µM). No additional interactions could be de-

tected as compared to the experiment done without the addition of ligand. This indicates the specificity of the ligand dependent interaction of CF16 with the estrogen receptor alpha.

EXAMPLE 5: DETECTION OF MUTANT ALLELES OF THE GENES ACCORDING TO THE INVENTION AND THEIR UTILISATION FOR DIAGNOSTIC PURPOSES.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type cofactor gene is detected. In addition, the method can be performed by detecting the wild-type cofactor gene and confirming the lack of cause of the disease as a result of the locus.

"Alteration of the wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and non-coding regions. Deletions may be of the entire gene or of only a portion. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues and are not inherited in the germline. Germline mutations can be found in any of a body's tissue and are mostly inherited. Point mutational events may occur in regulatory regions, such as the promotor of the gene, leading to loss or dimunition of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the cofactor gene product or to a decrease in mRNA stability or translation efficiency.

Applicable diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNAse protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, hybridization using nucleic acid modified with gold nanoparticles and PCR-SSCP, as discussed in detail further below. Furthermore, DNA microchip technology can be applied.

The presence of a disease due to a germline mutation of a cofactor can be ascertained by testing any tissue of the diseased human for mutations of the cofactor gene. For instance, a person who has inherited a germline mutation in the cofactor gene, especially one that will alter the interaction of the cofactor with the ER alpha protein, will be prone to develop a disease, such as cancer, bone diseases or defects in reproductive organs. The presence of such a mutation can be determined by extracting DNA from any tissue of the body. For example, blood can be drawn and DNA extracted from blood cells and analyzed. Moreover, prenatal

diagnosis of the disease will be possible by testing fetal cells, placental cells or amniotic cells for mutations in the cofactor gene. There are several methods that allow the detection of alterations of the wild-type cofactor gene, including for instance point mutations as well as deletions in the DNA sequence and these methods are discussed here:

Direct genomic DNA Sequencing, either manual or by automated means can detect sequence variations of cofactor genes (Nucleic Acids Res 1997 May 15;25(10):2032-2034 Direct DNA sequence determination from total genomic DNA. Kilger C, Pääbo S, Biol. Chem. 1997 Feb; 378(2):99-105, Direct exponential amplification and sequencing (DEXAS) of genomic DNA. Kilger C, Pääbo S, DE 19653439.9 and DE 19653494.1). Another way is to make use of the single-stranded conformation polymorphism assay (SSCP; Orita et al., PNAS 86, 2766 (1989)). Variations in the DNA sequence of the cofactor gene from the wild-type sequence will be detected due to a shifted mobility of the corresponding DNA-fragments in SSCP gels.

Other approaches are based on the detection of mismatches between the two complementary DNA strands. These methods, which will not allow the detection of large deletions, duplications or insertions nor the detection of a regulatory mutation affecting transcription or translation of the cofactor gene include the clamped denaturing gel electrophoresis (CDGE; Sheffield et al.,1991), heteroduplex analysis (HA; White et al., *Genomics* 4, 560 (1992)) and chemical mismatch cleavage (CMC; Grompe et al., 1989). Other methods detect specific types of mutations such as deletions, duplications or insertions, for instance a protein truncation assay or the asymmetric assay. These assay however, will not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a review by Grompe, *Nature Genetics* 5, 111 (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridisation will allow the rapid screening of a large number of other sample for that mutation. Such a technique may involve the utilisation of probes which are labeled with gold nanoparticles to yield a visual colour result (Elghanian et al., Science 277, 1078 (1997)).

In another embodiment of the present invention large scale genetic studies might be applied to investigate the association of a disease-phenotype with the gene of interest. The availability of the human genome allows an easy definition of genetic markers for most genes for a particular disease physiology. More importantly, single nucleotide polymorphisms (SNPs) are amenable markers for large genetic studies. SNPs in coding or regulatory regions of genes which

are thought to contribute to a disease physiology can have a direct impact on the phenotype, *e.g.* change a quantitative readout of disease physiology, for example the age of onset of heart attack. Association and linkage studies with related individuals, therefore provide an excellent means to test or verify a hypothesis on the functional impact of the gene of interest on disease physiology in vivo, in humans.

The ER alpha protein is known to control the expression of numerous estrogen responsive genes, which are implicated in the regulation of physiological and developmental processes such as sexual differentiation and behavior, fertility, cardiovascular function, brain function, bone generation and resorption as well as cell proliferation and carcinogenesis. Proteins interacting with ER alpha and/or the cofactors according to the invention are involved in the function of ER alpha. Therefore, alterations in the cofactors are useful for determining the genetic state of a person with respect to its capability to respond to estrogens.

In order to detect polymorphisms in DNA sequences, DNA samples can be prepared from normal individuals and from persons being affected by the disease and these samples can be cut by one or more restriction enzymes and applied to Southern analysis. Southern blots displaying hybridizing fragments differing in length from the control DNA when probed with sequences near or including the cofactor locus could indicate a possible mutation. If large DNA fragments are used it is appropriate to separate these fragments by pulsed field gel electrophoresis (PFGE).

Detection of point mutations may be accomplished by amplification, for instance by PCR, from genomic or cDNA and sequencing of the amplified nucleic or by molecular cloning of the cofactor allele and sequencing the allele using techniques well known in the art.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCP) (Orita et al., *PNAS*, 86, 2766 (1989)); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *NAR* 18, 2699, (1990); Sheffield et al., *PNAS* 86 , 232 (1989)); 3) RNase protection assays (Finkelstein et al., *Genomics* 7, 167 (1990); Kinszler et al., Science 251, 1366 (1991)); 4) allele specific oligonucleotides (ASOs, Conner et al., PNAS, 80, 278 (1983)); 5) the use of proteins which recognise nucleotide mismatches, such as the E. coli mutS protein (Modrich, *Ann. Rev. Genetics*, 25, 229 (1991)) and 6) allele-specific PCR (Ruano and Kidd, *NAR* 17,

8392 (1989)). For allele-specific PCR, primers are used which hybridise at their 3' ends to a particular *cofacto* mutation. Without the mutation, no PCR product is observed. Amplification Refractory Mutation System could also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., *NAR* 17, 2503 (1989). Insertions and deletions of genes can also be detected by molecular cloning, amplification and sequencing. Moreover, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score for alteration of an allele or an insertion in a polymorphic fragment. Such a method would be particularly useful for screening relatives of an affected person for the presence of the mutation found in that person. Other approaches for detecting insertions and deletions as known for those trained in the art can be used.

SSCP detects a band which migrates differently because the variation causes a difference in single strand, intra molecular base pairing. The RNAse protection assay involves cleavage of the mutant fragment into two or more smaller fragments. By using DGGE variations in the DNA can be detected by differences in the migration rates of mutant compared to normal alleles in a denaturing gradient gel. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a hetero duplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridised nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or the corresponding mRNA product. While these techniques are less sensitive than sequencing, they can preferably be used when a large number of samples shall be tested. An example of a mismatch cleavage method is the RNAse protection assay. In the practice of the present invention , the method involves the use of a labeled ribonucleotide probe which is complementary to the wild-type sequence of the cofactor gene coding sequence.The riboprobe and either mRNA or DNA isolated from the person are hybridised together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by the enzyme, it cleaves at the site of the mismatch. Consequently, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNAse A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. If the riboprobe comprises only a fragment of the mRNA or the gene,

it is advantageous to use a number of probes to screen the whole mRNA sequence for mismatches.

Similarly, DNA probes can be used to detect mismatch mutations through enzymatic or chemical cleavage (Cotton et al., *PNAS* 85, 4397 (1988); Shenk et al., *PNAS* 72, 989 (1975); Novack et al., *PNAS* 83, 586 (1986)). Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to match duplexes (Cariello, *Human Genetics* 42, 726 (1988)). With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridisation. Variations in DNA of the *cofactor)* gene can also be detected using Southern hybridisation, especially if the changes are major rearrangements, such as deletions or insertions. DNA sequences of the cofactor gene which have been amplified by PCR may also be screened using allele specific probes. These probes are nucleic acid oligomers, each of which contains a region of the gene sequence harboring a known mutation. For instance, one oligomer could be about 25 nucleotides in length corresponding to a portion of the gene sequence. By using a number of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously discovered mutation in the gene. Hybridisation of allele-specific probes with amplified *cofactor* sequences can be performed, for example, on a nylon filter. Under high stringency hybridisation conditions, the hybridisation of a particular probe should indicate the presence of the same mutation in the tissue as in the allele-specific probe.

The newly developed technique of nucleic acid analysis via microchip technology is also applicable to the present invention. In this technique, thousands of distinct nucleotide probes are built up in an array on a silicon chip. Nucleic acid to be analysed is fluorescently labeled and hybridised to the probes on the chip. It is also possible to study nucleic acid-protein interactions using these nucleic acid microchips. Using this technique one can determine the presence of mutations or even sequence the nucleic acid being analysed or one can measure expression of a gene of interest. This method is one of parallel processing of thousands of probes at once and can tremendously accelerate the analysis. In several publications the use of this method is described (Hacia et al., *Nature Genetics* 14, 441 (1996); Shoemaker et al., *Nature Genetics* 14, 450 (1996); Chee et al., *Science* 274, 610 (1996); DeRisi et al., *Nature Genetics* 14, 457 (1996)). This new technology has also been reviewed in Borman et al., *Chemi-*

cal and Engineering News 9, 42 (1996) and has been subject of an editorial in *Nature Genetics* (1996).

The most definite test for mutations in a candidate locus is to directly compare genomic cofactor sequences from patients with those from normal individuals. Alternatively one could sequence mRNA after amplification (for example by PCR) thereby eliminating the necessity of determining the exon structure of the respective gene.

Mutations from patients falling outside the coding region of the cofactor gene can be detected by examining the noncoding regions, such as introns and regulatory sequences within or near the genes. Early indications of mutations in noncoding regions could be for example the abundance or abnormal size of mRNA products in patients as compared to control individuals as detected by northern blot analysis.

Alteration of cofactor expression can be detected by any technique known in the art. These include northern blot analysis, PCR amplification and RNAse protection. Diminished mRNA expression indicates an alteration in the wild-type gene sequence. Alterations of wild-type genes can also be detected by screening for alteration of cofactor protein. For example, monoclonal antibodies against cofactor protein can be used to screen a tissue. Lack of cognate antigen would indicate a mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant gene product. These kind of immunological assays could be done in any convenient format known in the art. These include western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered cofactor protein can be used to detect alteration of the wild-type *cofactor* gene. Functional assays such as protein binding determinations can be used. Moreover, assays can be used which detect the cofactors' biochemical function. Finding a mutant cofactor gene product indicates an alteration of the cofactor wild-type gene. One such binding assay tests the binding of cofactor protein with wild-type ER alpha protein. Conversely, wild-type ER alpha protein or the domain interacting with the cofactor protein can be used in a protein binding assay or biochemical function assay to detect normal or mutant proteins.

A mutant cofactor gene or gene product or a mutant ER alpha protein can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant genes or gene products in tissues can be applied to other

body samples. By screening such body samples, a simple early diagnosis can be achieved for the disease) resulting from a mutation in the cofactor gene.

EXAMPLE 6: A CELL BASED ASSAY FOR MEASURING THE BINDING OF THE COFACTOR CF16, CF17, CF18, CF19, CF40, CF41, CF42 OR CF43, RESPECTIVELY TO ER ALPHA.

The DNA sequence encoding the open reading frame of the respective cofactor is transferred into the vector pVP16 (Clontech) to allow the expression of a fusion protein of the cofactor with the strong transactivation domain of the VP16 protein (of herpes simplex virus) in mammalian cells under the control of the strong CMV promoter. In another vector (the reporter), the luciferase gene is cloned under the control of a minimal promoter containing a ER alpha-responsive DNA element. This vector also expresses a second enzyme, e.g. beta-galactosidase, under the control of a constitutive promoter, to allow normalisation for transfection efficiency between experiments. A third vector contains the ER alpha gene under the control of the strong CMV promoter.

CV-1 cells are then transiently transfected with different combinations of the three plasmids. Transfection is done by standard methods, e.g. by use of the CalPhos Maximizer (Clontech, #8021-1,-2). Interaction of the cofactor protein with ER alpha will lead to a strong transactivation due to the attached VP16 domain of the cofactor fusion protein. Thus, interaction of the cofactor with ER alpha will result in increased luciferase activity. Inclusion of the cofactor VP16 will result in increased luciferase activity as compared to transfection of the ER alpha and the reporter alone. To measure this effect, extracts are prepared of the transfected cells 48 to 72 hours after transfection, and luciferase activity is determined. To normalise for transfection efficiency, beta-galactosidase activity is also determined.

Addition of substances known or suspected to influence the binding of ER alpha to the cofactor are added to the medium of the transfected cells. These substances are added at different timepoints prior to cell lysis, typically ranging between 18 hours to a five minutes before cell lysis. Luciferase activity is taken as a measure of the effect of these substances on the binding of the cofactor to ER alpha. To avoid activation of ER alpha by substances contained in the serum of the medium, charcoal stripped serum has to be used for these experiments.

In an alternative setting of the experiment, the DNA-binding domain of ER alpha is replaced with the DNA-binding domain of the yeast GAL4 transcription factor. On the reporter plasmid, the luciferase is expressed under the control of GAL4-responsive upstream activating sequences. Expression of luciferase again is an indication for binding of the cofactor CF16-VP16, CF17-VP16, CF18-VP16, CF19-VP16, CF40-VP16, CF41-VP16, CF42-VP16 or CF43-VP16, fusion respectively to the GAL4-ER alpha fusion. This setting is also referred to as the mammalian two hybrid system. A description of the experiment is found in the manual to the mammalian MATCHMAKER Two-Hybrid Assay Kit from Clontech, # PT3002-1, catalogue #K1602-1)

Substances activating nuclear receptors cause an exchange of the proteins bound to the receptors, thus effecting the dissociation of some proteins and promoting the binding of other proteins. Thus, in the experiments as described above, one can test for ER alpha-activating compounds and ER alpha-inactivating compounds by monitoring the binding of the respective CF cofactor to ER alpha.

In an alternative setting, stably transfected cell lines are used which contain copies of the two different expression constructs for ER alpha and the respective CF cofactor as well as the reporter construct stably integrated into the chromosomes of the cells.

EXAMPLE 7: A FRET ASSAY USING COFACTOR PROTEINS

DNA sequences encoding the open reading frame of the cofactor and the ER alpha gene are each transferred separately into the vector pENTRY (Life Technologies) to allow efficient construction of a diverse set of expression constructs. The open reading frame is then recombined into the vector pDEST17 for expression in E. coli strain BL21 as a fusion protein to a six-histidine tag induced by IPTG, as well as into the pDEST15 for expression as a fusion protein with glutathione S-transferase (GST). The plasmids pDEST15, pDEST17and pENTRY are purchased from LIFE TECHNOLOGIES. Alternatively, the open reading frame is introduced into the vector pLV-CBDgw for expression as a fusion protein with the calmodulin binding protein using recombinant baculoviruses as specified by the manufacturer (Becton Dickinson). pLV-CBDgw is a derivative of the vector pLV1392 (Becton Dickinson) which is modified by the insertion of a calmodulin binding protein fragment, followed by the sequence required for recombinational cloning via the Gateway system (Life Technologies). Protein

expression is induced and recombinant protein is purified by passage over a Ni-NTA-column, or a glutathione column or a calmodulin column, respectively.

To measure the interaction of the two proteins, a biotinylated (Biotintag Micro biotinylation Kit, Sigma) His-tagged ER alpha protein and the GST fusion of the cofactor are mixed at 0.2-5 μM. Antibody to the GST protein is added which is labelled by the europium chelate at a concentration of 1-3 (typical 2.5) nM. Streptavidin which is fluorescently labeled by covalent attachment of allophycocyanin is added at a concentration of 5-30 μg/ml (typical 10μg/ml). The europium chelate is stimulated by a flash of light (320nm) and, the emitted light is measured in a delayed (50-200 μs) time window for 300 to 450 μs after the flash at 615 nm (fluorescence of europium chelate) and 655nm (fluorescence of APC). Since APC is only excited by the light emitted by the europium chelate, a close proximity of the two different fluorophores is required for excitation. The strength of the APC signal, as well as the ratio of the signals from the two fluorophores (i.e. the ratio of the intensities of light emitted at 655 and 615nm) serves as a measure for the interaction of the two proteins. Reaction buffers contain 20mM TrisHCl pH 7.9, 60mM KCl, 4mM $MgCl_2$. Reaction volume is 25μl. The Wallac VictorV fluorimeter is used for the fluorimetric measurements.

In an alternative setting, the cofactor is used as a biotinylated His-tagged protein, and the ER alpha protein is used as fusion to GST. In yet another setting, the His-tagged proteins are replaced by the same proteins fused to the calmodulin binding protein. In the latter case, the detection of the interaction is via biotinylated calmodulin, which is in turn binding to APC-coupled streptavidin. Calcium has to be included in the buffer in the form of 4mM $CaCl_2$, to allow complex formation between calmodulin and the calmodulin binding protein.

CLAIMS:

1. An isolated nucleic acid molecule coding for a cofactor (CF) of the estrogen receptor alpha which is selected from the group consisting of:

   a) the nucleotide sequences set forth in SEQ ID NOs: 1, 4, 7, 10, 13, 16, 19, and/or 22;

   b) or complements thereof as set forth in SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, and/or 23;

   c) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs: 1, , 4, 7, 10, 13, 16, 19, and/or 22; under conditions of high stringency, and

   d) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, and/or 23, under conditions of high stringency.

2. The isolated nucleic acid molecule of claim 1 which is genomic DNA.

3. The isolated nucleic acid molecule of claim 1 which is cDNA.

4. The isolated nucleic acid molecule of claim 1 which is RNA.

5. An isolated nucleic acid molecule comprising the nucleic acid molecule of any of claims 1 to 4 and a label attached thereto.

6. A vector comprising the nucleic acid molecule of claim 1.

7. The vector of claim 6, which is an expression vector.

8. A host cell transfected with the vector of claim 6 or 7.

9. A host cell transfected with the expression vector of claim 7.

10. A method of producing a polypeptide comprising the step of culturing the host cell of claim 9 in an appropriate culture medium to thereby produce the polypeptide.

11. An isolated polypeptide encoded by any portion of a nucleic acid of claim 1.

12. An isolated polypeptide selected from the group consisting of the amino acid sequences set forth in SEQ ID NOs.: 3, 6, 9, 12, 15, 18, 21, and/or 24.

13. Proteinous complex, comprising a cofactor polypeptide according to any of SEQ ID NOs.: 3, 6, 9, 12, 15, 18, 21, and/or 24 or a portion thereof together with a estrogen receptor alpha polypeptide according to any of SEQ ID NOs. 27 or 30 or a portion thereof.

14. Complex according to claim 13, further comprising at least one non-protein cofactor.

15. A method for screening for compounds which are capable of inhibiting the cellular function of at least one of the cofactors CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43, in particular binding of said cofactors to estrogen receptor alpha polypeptide according to SEQ ID NOs. 27 or 30 or a portion thereof comprising the steps of:

    a) contacting one or more candidate compounds with a polypeptide according to claims 11, 12 or a complex according to any of claims 13 or 14,
    b) removing unbound compound(s),
    c) detecting whether the compounds(s) interact with the polypeptide of the cofactor.

16. A method for screening for compounds which are capable of inhibiting or activating the cellular function of estrogen receptor alpha polypeptide according to SEQ ID NOs. 27 or 30 or a portion thereof, in particular binding of said liver X receptor alpha polypeptide to at least one of the cofactors CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43 comprising the steps of:

    a) contacting one or more candidate compounds which are capable of binding with a complex according to any of claims 13 to 15,
    b) removing unbound compound(s),

c) detecting the amount of the polypeptide of the CF according to any of claims 11 or 12 of that remained bound within the complex and

d) identifying such compounds capable of either: i) releasing a large amount of the polypeptide of the CF according to any of claims 11 or 12 from the complex, or ii) promoting the association of CF polypeptides according to any of claims 11 or 12 from the complex.

17. Compound or mixture of compounds, identified by the method according to claim 16.

18. A method for inhibiting or activating the cellular function of the cofactor CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43, comprising the steps of:

a) contacting a cell with a binding agent that binds the polypeptide according to claim 11, 12 or the complex according to any of claims 13 or 145,

b) whereby the cellular function of CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43 is inhibited or activated.

19. A method for inhibiting or activating the binding of the cofactor CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43 to an estrogen receptor alpha polypeptide according to SEQ ID NOs. 27 or 30 or a portion thereof comprising the steps of:

a) contacting the polypeptide according to claim 11, 12 or the complex according to claim 13 or 14 with a binding agent,

b) whereby the binding of CF16, CF17, CF18, CF19, CF40, CF41, CF42 and/or CF43 to the estrogen receptor alpha polypeptide is inhibited or activated.

20. Method according to claim 18 or 19, characterized in that the binding agent is an antibody.

21. Method according to claim 18 or 19, characterized in that the binding agent is RNA.

22. Method according to claim 18 or 19, characterized in that the binding agent is an antisense oligonucleotide.

23. Method according to claim 18 or 19, characterized in that the binding agent is a ribozyme.

24. Method according to claim 18 or 19, characterized in that the binding agent is a steroid molecule.

25. Method according to claim 18 or 19, characterized in that the cell is in a body.

Figure 1A

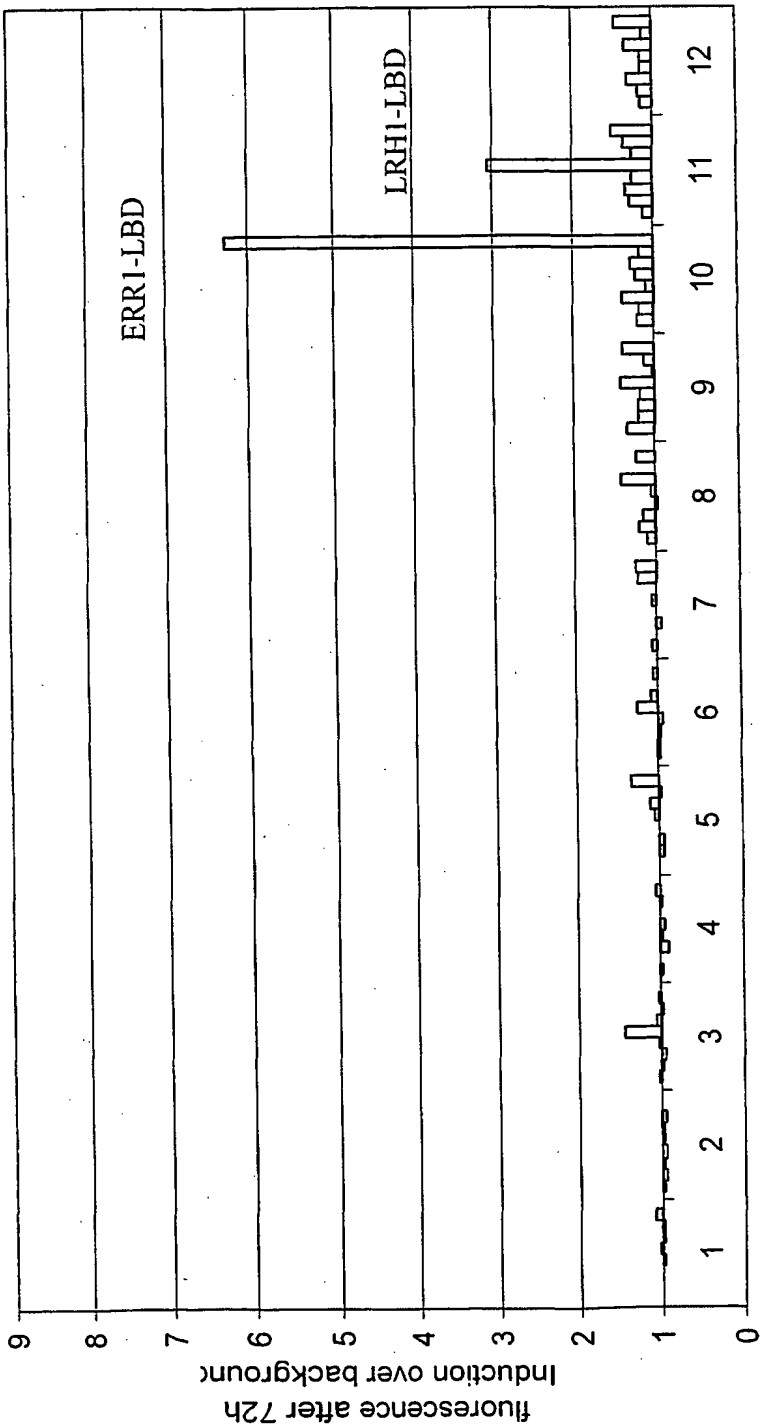
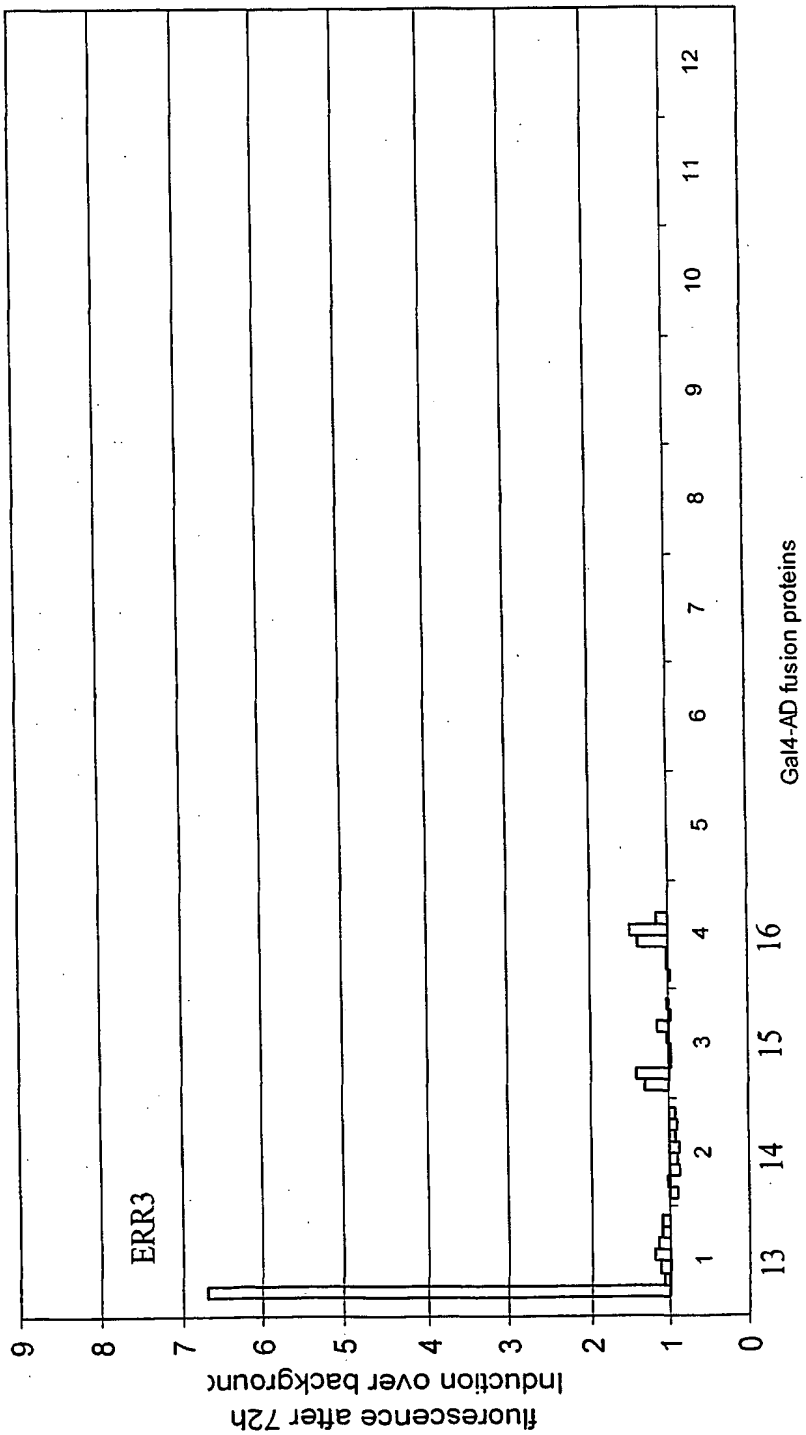


Figure 1A (continued)



no ligand

Figure 1B

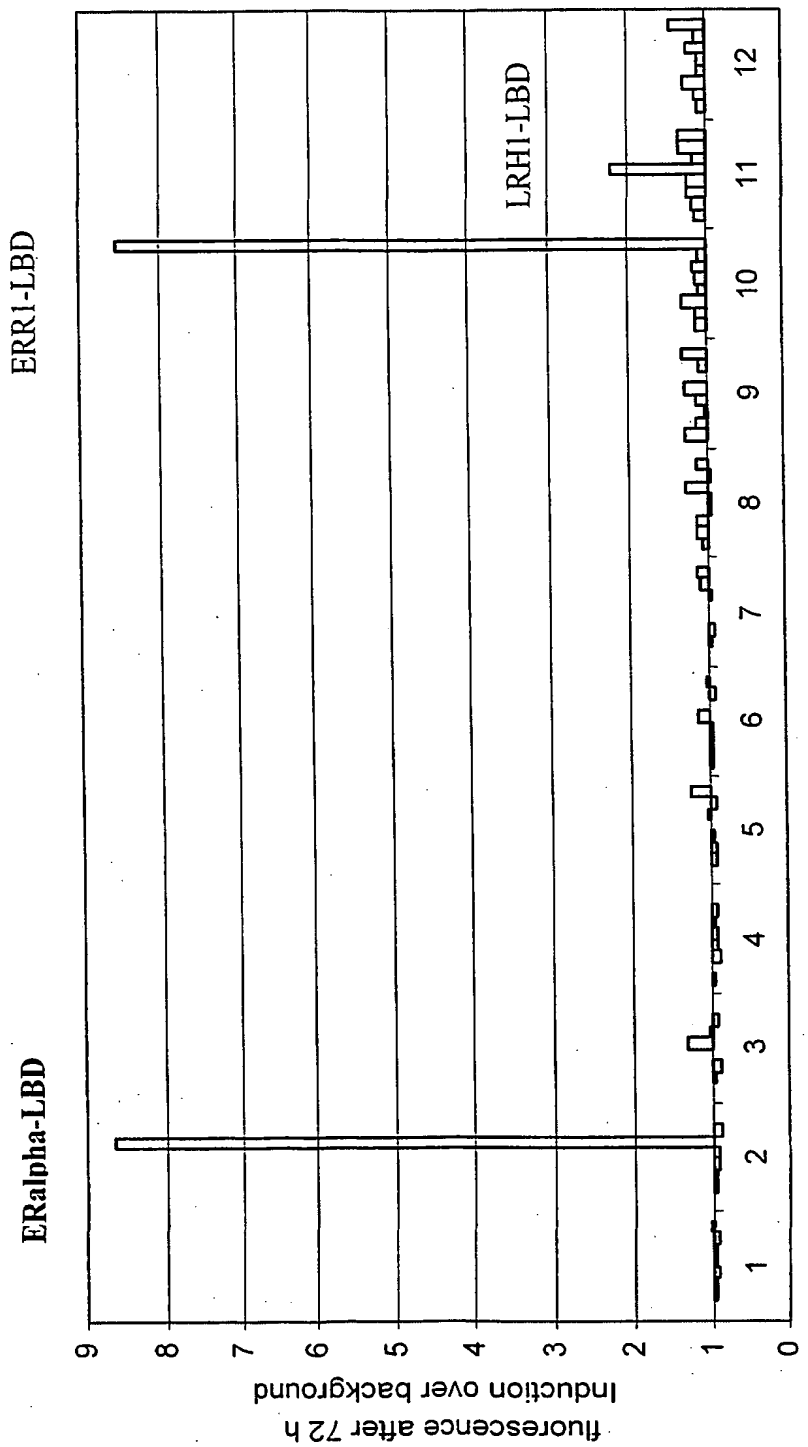
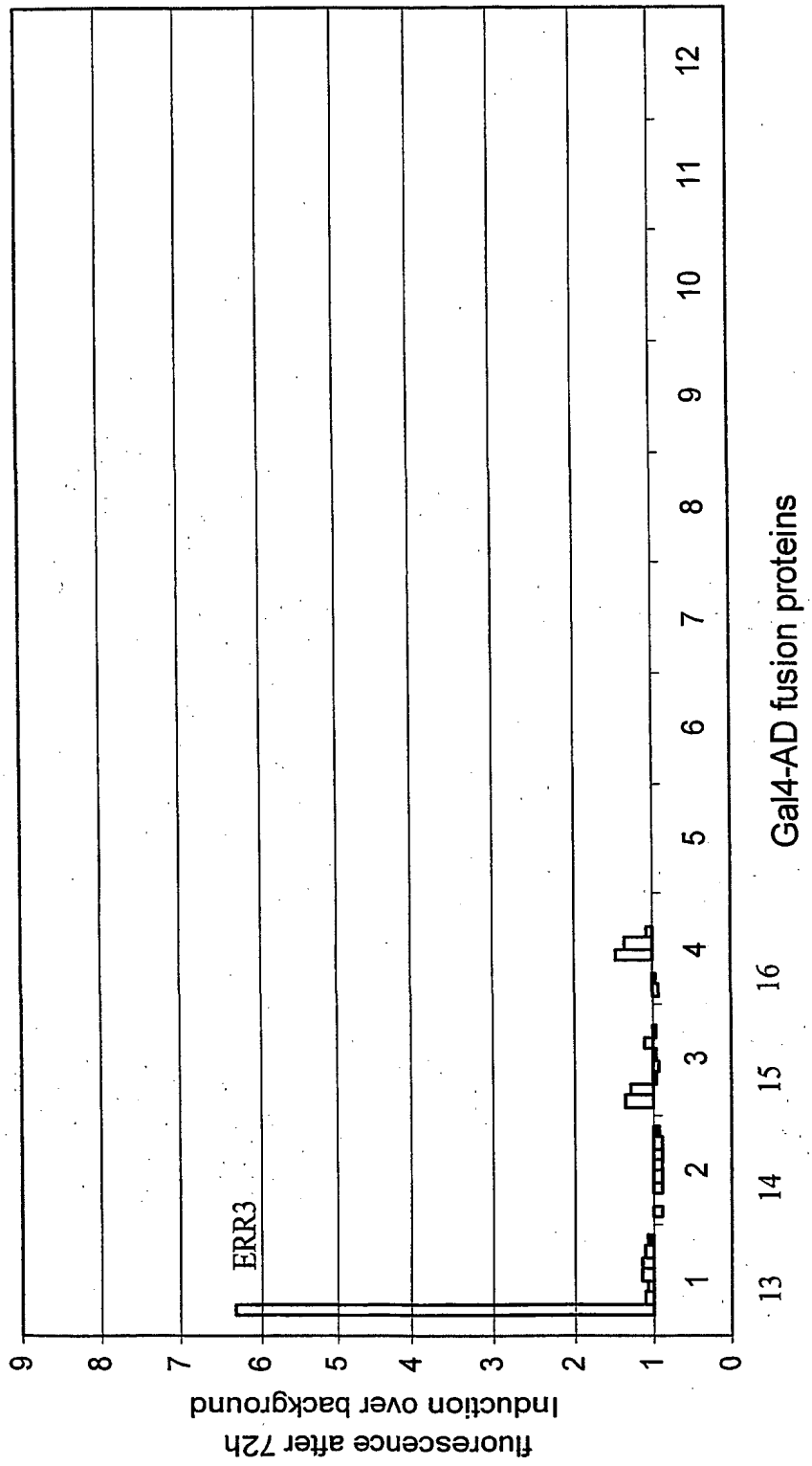

ERalpha-LBD
ERR1-LBD
LRH1-LBD
fluorescence after 72 h
induction over background
0
1
2
3
4
5
6
7
8
9
1
2
3
4
5
6
7
8
9
10
11
12
Gal4-AD fusion proteins
estradiol 250 nM

Figure 1B (continued)



SEQUENCE LISTING

<110> LION bioscience AG
<120> Novel Cofactors of the Estrogen Receptor alpha and methods of use
<130> L30202PCT
<160> 30

<170> PatentIn version 3.1

<210> 1
<211> 2133
<212> DNA
<213> Homo sapiens

<400> 1
agctggctgg gcggttagga gggcccgggg ccgagacgat ggctgaccac aaccctgaca 60

gcgactccac gccgcgcacg ctgctgcgac gcgtgctgga tacagcggac ccgcgcaccc 120

cgcggcgacc ccggagtgct cgggctggag cccggagagc cctgcttgaa acggcttccc 180

ccaggaagtt gagtggccaa acaaggacga tagccagagg gcgttcccat ggagccaggt 240

ctgttggcag atcggcccat attcaggcca gtgggcactt ggaggaacag acacctcgga 300

cgctgctgaa gaacatccta ctaactgccc cagaatcttc catcctgatg cctgagtcgg 360

tagtgaagcc agtgccagca ccgcaggcgg tccaaccctc cagacaagag agcagttgcg 420

gcagcctgga gctgcaactt cctgagctcg agccccccac aaccctggct ccaggtctgc 480

tggcccctgg caggaggaaa cagaggctga gactgtcagt gtttcagcag ggagtggacc 540

aggggctgtc tctctcccaa gagcctcaag ggaatgctga tgcctcttcc ctcaccagat 600

ccctcaacct gacctttgcc acgcctcttc agccacagtc agtgcagagg cctggcttgg 660

cccgcagacc tccagcccgc cgagctgtag acgtgggtgc ctttttgcgg gatctgcgag 720

atacttccct ggctcctcca aacattgtgt tggaggacac ccagccgttc tctcagccca 780

tggttggctc ccccaacgtg tatcactccc tgccctgcac gcctcacact ggggctgaag 840

acgctgagca ggctgccggt cgcaagacac agagcagtgg gcctgggctg cagaagaata 900

gtgagtgtgt ggcactggtg gcctggagcc aaatttagct tgggtgagag ttgacaatgg 960

tagttttcct tcctcaagcc cctctgtgcc cctagagcac cctggctgtg gctgcctcct 1020

tcatccaaga gcagagtcca tgttgggcca ggagacttca gatccatgtc ctggtgctgc 1080

ctctggcttt gtctttcctc agtgggcagg actgggtctg ctggtccatc tttacccttc 1140

tctgagctat gcagccttgg cctgctgcgt ctccggcctg tattctctcc ccttcactca 1200

ggccctggga aaccagccca gtttctggca ggagaggcag aggaggtcaa tgcctttgct 1260

ctgggcttcc tgagcaccag cagtggtgtc tctggagaag atgaagtaga gcccttacac 1320

gatggagttg aagaggcaga gaaaaagatg gaagaagaag gtgtgagtgt gagtgaaatg 1380

gaggcaacag gagcacaagg acccagcagg gtagaagagg ctgagggaca cacagaggtg 1440

acagaagcag agggatccca ggggactgct gaggctgacg ggccaggagc atcttcaggg 1500

gatgaggatg cctctggcag ggcagcaagt ccagagtcgg cctccagcac ccctgagtct 1560

ctccaggcca ggcgacatca tcagtttctt gagccagccc cagcgcctgg tgctgcagtc 1620

ttatcttcag agcctgcaga gcctctgttg gtcaggcatc cccctaggcc ccggaccacc 1680

ggccccaggc cccggcaaga tccccacaag gctggactga gccactatgt gaaactcttt 1740

agcttctatg ccaagatgcc catggagagg aaggctcttg agatggtgga gaagtgccta 1800

gataaatatt tccagcatct ttgtgatgat ctggaggtat ttgctgctca tgctggccgc 1860

aagactgtga agccagagga cctggagctg ctgatgcggc ggcagggcct ggtcactgac 1920

caagtctcac tgcacgtgct agtggagcgg cacctgcccc tggagtaccg gcagctgctc 1980

atcccctgtg catacagtgg caactctgtc ttccctgccc agtagtggcc aggcttcaac 2040

actttccctg tccccacctg gggactcttg cccccacata tttctccagg tctcctcccc 2100

acccccccag catcaataaa gtgtcataaa cag 2133

<210> 2
<211> 2133
<212> DNA
<213> Homo sapiens

<400> 2
ctgtttatga cactttattg atgctggggg ggtggggagg agacctggag aaatatgtgg 60

gggcaagagt ccccaggtgg ggacagggaa agtgttgaag cctggccact actgggcagg 120

gaagacagag ttgccactgt atgcacaggg gatgagcagc tgccggtact ccaggggcag 180

gtgccgctcc actagcacgt gcagtgagac ttggtcagtg accaggccct gccgccgcat 240

cagcagctcc aggtcctctg gcttcacagt cttgcggcca gcatgagcag caaatacctc 300

cagatcatca caaagatgct ggaaatattt atctaggcac ttctccacca tctcaagagc 360

cttcctctcc atgggcatct tggcatagaa gctaaagagt ttcacatagt ggctcagtcc 420

agccttgtgg ggatcttgcc ggggcctggg gccggtggtc cggggcctag ggggatgcct 480

gaccaacaga ggctctgcag gctctgaaga taagactgca gcaccaggcg ctggggctgg 540

ctcaagaaac tgatgatgtc gcctggcctg gagagactca ggggtgctgg aggccgactc 600

tggacttgct gccctgccag aggcatcctc atcccctgaa gatgctcctg gcccgtcagc 660

ctcagcagtc ccctgggatc cctctgcttc tgtcacctct gtgtgtccct cagcctcttc 720

taccctgctg ggtccttgtg ctcctgttgc ctccatttca ctcacactca caccttcttc 780

ttccatcttt ttctctgcct cttcaactcc atcgtgtaag ggctctactt catcttctcc 840

agagacacca ctgctggtgc tcaggaagcc cagagcaaag gcattgacct cctctgcctc 900

tcctgccaga aactgggctg gtttcccagg gcctgagtga aggggagaga atacaggccg 960

gagacgcagc aggccaaggc tgcatagctc agagaagggt aaagatggac cagcagaccc 1020

agtcctgccc actgaggaaa gacaaagcca gaggcagcac caggacatgg atctgaagtc 1080

tcctggccca acatggactc tgctcttgga tgaaggaggc agccacagcc agggtgctct 1140

aggggcacag aggggcttga ggaaggaaaa ctaccattgt caactctcac ccaagctaaa 1200

tttggctcca ggccaccagt gccacacact cactattctt ctgcagccca ggcccactgc 1260

tctgtgtctt gcgaccggca gcctgctcag cgtcttcagc cccagtgtga ggcgtgcagg 1320

gcagggagtg atacacgttg ggggagccaa ccatgggctg agagaacggc tgggtgtcct 1380

ccaacacaat gtttggagga gccagggaag tatctcgcag atcccgcaaa aaggcaccca 1440

cgtctacagc tcggcgggct ggaggtctgc gggccaagcc aggcctctgc actgactgtg 1500

gctgaagagg cgtggcaaag gtcaggttga gggatctggt gagggaagag gcatcagcat 1560

tcccttgagg ctcttgggag agagacagcc cctggtccac tccctgctga aacactgaca 1620

gtctcagcct ctgtttcctc ctgccagggg ccagcagacc tggagccagg gttgtggggg 1680

gctcgagctc aggaagttgc agctccaggc tgccgcaact gctctcttgt ctggagggtt 1740

ggaccgcctg cggtgctggc actggcttca ctaccgactc aggcatcagg atggaagatt 1800

ctggggcagt tagtaggatg ttcttcagca gcgtccgagg tgtctgttcc tccaagtgcc 1860

cactggcctg aatatgggcc gatctgccaa cagacctggc tccatgggaa cgccctctgg 1920

ctatcgtcct tgtttggcca ctcaacttcc tgggggaagc cgtttcaagc agggctctcc 1980

gggctccagc ccgagcactc cggggtcgcc gcggggtgcg cgggtccgct gtatccagca 2040

cgcgtcgcag cagcgtgcgc ggcgtggagt cgctgtcagg gttgtggtca gccatcgtct 2100

cggccccggg ccctcctaac cgcccagcca gct 2133

<210> 3
<211> 299
<212> PRT
<213> Homo sapiens

<400> 3

Met Ala Asp His Asn Pro Asp Ser Asp Ser Thr Pro Arg Thr Leu Leu
1 5 10 15

Arg Arg Val Leu Asp Thr Ala Asp Pro Arg Thr Pro Arg Arg Pro Arg
20 25 30

Ser Ala Arg Ala Gly Ala Arg Arg Ala Leu Leu Glu Thr Ala Ser Pro
35 40 45

Arg Lys Leu Ser Gly Gln Thr Arg Thr Ile Ala Arg Gly Arg Ser His
50 55 60

Gly Ala Arg Ser Val Gly Arg Ser Ala His Ile Gln Ala Ser Gly His
65 70 75 80

Leu Glu Glu Gln Thr Pro Arg Thr Leu Leu Lys Asn Ile Leu Leu Thr
85 90 95

Ala Pro Glu Ser Ser Ile Leu Met Pro Glu Ser Val Val Lys Pro Val
100 105 110

Pro Ala Pro Gln Ala Val Gln Pro Ser Arg Gln Glu Ser Ser Cys Gly
115 120 125

Ser Leu Glu Leu Gln Leu Pro Glu Leu Glu Pro Pro Thr Thr Leu Ala
130 135 140

Pro Gly Leu Leu Ala Pro Gly Arg Arg Lys Gln Arg Leu Arg Leu Ser
145 150 155 160

Val Phe Gln Gln Gly Val Asp Gln Gly Leu Ser Leu Ser Gln Glu Pro
165 170 175

Gln Gly Asn Ala Asp Ala Ser Ser Leu Thr Arg Ser Leu Asn Leu Thr
180 185 190

Phe Ala Thr Pro Leu Gln Pro Gln Ser Val Gln Arg Pro Gly Leu Ala
195 200 205

Arg Arg Pro Pro Ala Arg Arg Ala Val Asp Val Gly Ala Phe Leu Arg
210 215 220

Asp Leu Arg Asp Thr Ser Leu Ala Pro Pro Asn Ile Val Leu Glu Asp
225 230 235 240

Thr Gln Pro Phe Ser Gln Pro Met Val Gly Ser Pro Asn Val Tyr His
245 250 255

Ser Leu Pro Cys Thr Pro His Thr Gly Ala Glu Asp Ala Glu Gln Ala
260 265 270

Ala Gly Arg Lys Thr Gln Ser Ser Gly Pro Gly Leu Gln Lys Asn Ser
275 280 285

Glu Cys Val Ala Leu Val Ala Trp Ser Gln Ile
290 295

<210> 4
<211> 445

<212> DNA
<213> Homo sapiens

<400> 4
gccttccctt tgcaccagcc cgagtctagg tctgggccaa gcacattaca agtgggaccg 60

gtggagcagc ccctgggctc cctgggcagg ggagttctga ggctcctgct ctcccatcca 120

cctgtctgtc ctggcctaat gccaggctct gagttctgtg accaaagcca ggtgggttcc 180

ctttccttcc caccctgtg gccacagctc tggagtggga gggttggttg cccctcacct 240

cagagctccc ccaaaggcca gtaatggatc cccggcctca gtccctactc tgctttggga 300

tagtgtgagc ttcattttgt acacgtgtga cttcgtccag ttacaaaccc aataaactct 360

gtagagtgga aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 420

aaaaaaaaaa aaaaccccga aaaaa 445

<210> 5
<211> 445
<212> DNA
<213> Homo sapiens

<400> 5
ttttttcggg gttttttttt ttttttttt ttttttttt ttttttttt ttttttttt 60

ttttttttt ttttttccac tctacagagt ttattgggtt tgtaactgga cgaagtcaca 120

cgtgtacaaa atgaagctca cactatccca aagcagagta gggactgagg ccggggatcc 180

attactggcc tttgggggag ctctgaggtg aggggcaacc aaccctccca ctccagagct 240

gtggccacag gggtgggaag gaaagggaac ccacctggct ttggtcacag aactcagagc 300

ctggcattag gccaggacag acaggtggat gggagagcag gagcctcaga actcccctgc 360

ccagggagcc caggggctgc tccaccggtc ccacttgtaa tgtgcttggc ccagacctag 420

actcgggctg gtgcaaaggg aaggc 445

<210> 6
<211> 87
<212> PRT
<213> Homo sapiens

<400> 6

Ala Phe Pro Leu His Gln Pro Glu Ser Arg Ser Gly Pro Ser Thr Leu
1 5 10 15

Gln Val Gly Pro Val Glu Gln Pro Leu Gly Ser Leu Gly Arg Gly Val
20 25 30

Leu Arg Leu Leu Leu Ser His Pro Pro Val Cys Pro Gly Leu Met Pro
35 40 45

Gly Ser Glu Phe Cys Asp Gln Ser Gln Val Gly Ser Leu Ser Phe Pro
50 55 60

Pro Leu Trp Pro Gln Leu Trp Ser Gly Arg Val Gly Cys Pro Ser Pro
65 70 75 80

Gln Ser Ser Pro Lys Gly Gln
85

<210> 7
<211> 1234
<212> DNA
<213> Homo sapiens

<400> 7
atgtcgctga ccaacacaaa gacggggttt tcggtcaagg acatcttaga cctgccggac 60

accaacgatg aggagggctc tgtggccgaa ggtccggagg aagagaacga ggggcccgag 120

ccagccaaga gggccgggcc gctggggcag ggcgccctgg acgcggtgca gagcctgccc 180

ctgaagaacc ccttctacga cagcagcgac aacccgtaca cgcgctggct ggccagcacc 240

gagggccttc agtactccct gcacggtctg gctgccgggg cgccccctca ggactcaagc 300

tccaagtccc cggagccctc ggccgacgag tcaccggaca atgacaagga gaccccgggc 360

ggcgggggcg acgccggcaa gaagcgaaag cggcgagtgc ttttctccaa ggcgcagacc 420

tacgagctgg agcggcgctt tcggcagcag cggtacctgt cggcgcccga gcgcgaacac 480

ctggccagcc tcatccgcct cacgcccacg caggtcaaga tctggttcca gaaccaccgc 540

tacaagatga agcgcgcccg ggccgagaaa ggtatggagg tgacgccccc gccctcgccg 600

cgccgggtgg ccgtgcccgt cttggtcagg gacggcaaac catgtcacgc gctcaaagcc 660

caggacctgg cagccgccac cttccaggcg ggcattccct tttctgccta cagcgcgcag 720

tcgctgcagc acatgcagta caacgcccag tacagctcgg ccagcacccc ccagtacccg 780

acagcacacc ccctggtcca ggcccagcag tggacttggt gagcgccgcc ccaacgagac 840

tcgcggcccc aggcccaggc cccaccccgg cggcggtggc ggcgaggagg cctcggtcct 900

tatggtggtt attattatta ttataattat tattatggag tcgagttgac tctcggctcc 960

actagggagg cgccgggagg ttgcctgcgt ctccttggag tggcagattc cacccaccca 1020

gctctgccca tgcctctcct tctgaacctt gggagagggc tgaactctac gccgtgttta 1080

cagaatgttt gcgcagcttc gcttctttgc ctctccccgg ggggaccaaa ccgtcccagc 1140

gttaatgtcg tcacttgaaa acgagaaaaa gaccgacccc ccacccctgc tttcgtgcat 1200

tttgtaaaat atgtttgtgt gagtagcgat attg 1234

<210> 8
<211> 1234

<212> DNA
<213> Homo sapiens

<400> 8
caatatcgct actcacacaa acatatttta caaaatgcac gaaagcaggg gtgggggtc 60

ggtctttttc tcgttttcaa gtgacgacat taacgctggg acggtttggt ccccccgggg 120

agaggcaaag aagcgaagct gcgcaaacat tctgtaaaca cggcgtagag ttcagccctc 180

tcccaaggtt cagaaggaga ggcatgggca gagctgggtg ggtggaatct gccactccaa 240

ggagacgcag gcaacctccc ggcgcctccc tagtggagcc gagagtcaac tcgactccat 300

aataataatt ataataataa taataaccac cataaggacc gaggcctcct cgccgccacc 360

gccgccgggg tggggcctgg gcctggggcc gcgagtctcg ttggggcggc gctcaccaag 420

tccactgctg ggcctggacc agggggtgtg ctgtcgggta ctgggggtg ctggccgagc 480

tgtactgggc gttgtactgc atgtgctgca gcgactgcgc gctgtaggca gaaaagggaa 540

tgcccgcctg gaaggtggcg gctgccaggt cctgggcttt gagcgcgtga catggtttgc 600

cgtccctgac caagacgggc acggccaccc ggcgcggcga gggcaggggc gtcacctcca 660

tacctttctc ggcccgggcg cgcttcatct tgtagcggtg gttctggaac cagatcttga 720

cctgcgtggg cgtgaggcgg atgaggctgg ccaggtgttc gcgctcgggc gccgacaggt 780

accgctgctg ccgaaagcgc cgctccagct cgtaggtctg cgccttggag aaaagcactc 840

gccgctttcg cttcttgccg gcgtcccccc cgccgcccgg ggtctccttg tcattgtccg 900

gtgactcgtc ggccgagggc tccggggact tggagcttga gtcctgaggg ggcgccccgg 960

cagccagacc gtgcagggag tactgaaggc cctcggtgct ggccagccag cgcgtgtacg 1020

ggttgtcgct gctgtcgtag aaggggttct tcaggggcag gctctgcacc gcgtccaggg 1080

cgccctgccc cagcggcccg gccctcttgg ctggctcggg cccctcgttc tcttcctccg 1140

gaccttcggc cacagagccc tcctcatcgt tggtgtccgg caggtctaag atgtccttga 1200

ccgaaaaccc cgtctttgtg ttggtcagcg acat 1234

<210> 9
<211> 273
<212> PRT
<213> Homo sapiens

<400> 9

Met Ser Leu Thr Asn Thr Lys Thr Gly Phe Ser Val Lys Asp Ile Leu
1 5 10 15

Asp Leu Pro Asp Thr Asn Asp Glu Glu Gly Ser Val Ala Glu Gly Pro
20 25 30

Glu Glu Glu Asn Glu Gly Pro Glu Pro Ala Lys Arg Ala Gly Pro Leu

35 40 45

Gly Gln Gly Ala Leu Asp Ala Val Gln Ser Leu Pro Leu Lys Asn Pro
50 55 60

Phe Tyr Asp Ser Ser Asp Asn Pro Tyr Thr Arg Trp Leu Ala Ser Thr
65 70 75 80

Glu Gly Leu Gln Tyr Ser Leu His Gly Leu Ala Ala Gly Ala Pro Pro
85 90 95

Gln Asp Ser Ser Ser Lys Ser Pro Glu Pro Ser Ala Asp Glu Ser Pro
100 105 110

Asp Asn Asp Lys Glu Thr Pro Gly Gly Gly Gly Asp Ala Gly Lys Lys
115 120 125

Arg Lys Arg Arg Val Leu Phe Ser Lys Ala Gln Thr Tyr Glu Leu Glu
130 135 140

Arg Arg Phe Arg Gln Gln Arg Tyr Leu Ser Ala Pro Glu Arg Glu His
145 150 155 160

Leu Ala Ser Leu Ile Arg Leu Thr Pro Thr Gln Val Lys Ile Trp Phe
165 170 175

Gln Asn His Arg Tyr Lys Met Lys Arg Ala Arg Ala Glu Lys Gly Met
180 185 190

Glu Val Thr Pro Leu Pro Ser Pro Arg Arg Val Ala Val Pro Val Leu
195 200 205

Val Arg Asp Gly Lys Pro Cys His Ala Leu Lys Ala Gln Asp Leu Ala
210 215 220

Ala Ala Thr Phe Gln Ala Gly Ile Pro Phe Ser Ala Tyr Ser Ala Gln
225 230 235 240

Ser Leu Gln His Met Gln Tyr Asn Ala Gln Tyr Ser Ser Ala Ser Thr
245 250 255

Pro Gln Tyr Pro Thr Ala His Pro Leu Val Gln Ala Gln Gln Trp Thr
260 265 270

Trp

<210> 10
<211> 1727
<212> DNA
<213> Homo sapiens

<400> 10
gccatggcca ccacggtgac ctgcacccgc ttcaccgacg agtaccagct ctacgaggat 60
attggcaagg gggctttctc tgtggtccga cgctgtgtca agctctgcac cggccatgag 120
tatgcagcca agatcatcaa caccaagaag ctgtcagcca gagatcacca gaagctggag 180
agagaggctc ggatctgccg ccttctgaag cattccaaca tcgtgcgtct ccacgacagc 240
atctccgagg agggcttcca ctacctggtc ttcgatctgg tcactggtgg ggagctcttt 300
gaagacattg tggcgagaga gtactacagc gaggctgatg ccagtcactg tatccagcag 360
atcctggagg ccgttctcca ttgtcaccaa atgggggtcg tccacagaga cctcaagccg 420
gagaacctgc ttctggccag caagtgcaaa ggggctgcag tgaagctggc agacttcggc 480
ctagctatcg aggtgcaggg ggaccagcag gcatggtttg gtttcgctgg cacaccaggc 540
tacctgtccc ctgaggtcct tcgcaaagag gcgtatggca agcctgtgga catctgggca 600
tgtggggtga tcctgtacat cctgctcgtg ggctacccac ccttctggga cgaggaccag 660
cacaagctgt accagcagat caaggctggt gcctatgact tcccgtcccc tgagtgggac 720
accgtcactc ctgaagccaa aaacctcatc aaccagatgc tgaccatcaa ccctgccaag 780
cgcatcacag cccatgaggc cctgaagcac ccgtgggtct gccaacgctc cacggtagca 840
tccatgatgc acagacagga gactgtggag tgtctgaaaa agttcaatgc caggagaaag 900
ctcaagggag ccatcctcac caccatgctg gccacacgga atttctcagt gggcagacag 960
accaccgctc cggccacaat gtccaccgcg gcctccggca ccaccatggg gctggtggaa 1020
caagccaaga gtttactcaa caagaaagca gatggagtca agccccagac gaatagcacc 1080
aaaaacagtg cagccgccac cagccccaaa gggacgcttc ctcctgccgc cctggagcct 1140
caaaccaccg tcatccataa cccagtggac gggattaagg agtcttctga cagtgccaat 1200
accaccatag aggatgaaga cgctaaagcc cggaagcagg agatcattaa gaccacggag 1260
cagctcatcg aggccgtcaa caacggtgac tttgaggcct acgcgaaaat ctgtgaccca 1320
gggctgacct cgtttgagcc tgaagcactg ggcaacctgg ttgaagggat ggacttccac 1380
agattctact tcgagaacct gctggccaag aacagcaagc cgatccacac gaccatcctg 1440
aacccacacg tgcacgtcat tggagaggat gccgcctgca tcgcttacat ccggctcacg 1500
cagtacattg acgggcaggg ccggccccgc accagccagt ctgaggagac ccgcgtgtgg 1560
caccgccgcg acggcaagtg gcagaacgtg cacttccact gctcgggcgc gcctgtggcc 1620
ccgctgcagt gaagagctgc gccctggttt cgccggacag agttggtgtt tggagcccga 1680
ctgccctcgg gcacacggcc tgcctgtcgc atgtttgtgt ctgcctc 1727

<210> 11
<211> 1727
<212> DNA
<213> Homo sapiens

<400> 11
gaggcagaca caaacatgcg acaggcaggc cgtgtgcccg agggcagtcg ggctccaaac 60

accaactctg tccggcgaaa ccagggcgca gctcttcact gcagcggggc cacaggcgcg 120

cccgagcagt ggaagtgcac gttctgccac ttgccgtcgc ggcggtgcca cacgcgggtc 180

tcctcagact ggctggtgcg gggccggccc tgcccgtcaa tgtactgcgt gagccggatg 240

taagcgatgc aggcggcatc ctctccaatg acgtgcacgt gtgggttcag gatggtcgtg 300

tggatcggct tgctgttctt ggccagcagg ttctcgaagt agaatctgtg gaagtccatc 360

ccttcaacca ggttgcccag tgcttcaggc tcaaacgagg tcagccctgg gtcacagatt 420

ttcgcgtagg cctcaaagtc accgttgttg acggcctcga tgagctgctc cgtggtctta 480

atgatctcct gcttccgggc tttagcgtct tcatcctcta tggtggtatt ggcactgtca 540

gaagactcct taatcccgtc cactgggtta tggatgacgg tggtttgagg ctccagggcg 600

gcaggaggaa gcgtcccttt ggggctggtg gcggctgcac tgtttttggt gctattcgtc 660

tggggcttga ctccatctgc tttcttgttg agtaaactct tggcttgttc caccagcccc 720

atggtggtgc cggaggccgc ggtggacatt gtggccggag cggtggtctg tctgcccact 780

gagaaattcc gtgtggccag catggtggtg aggatggctc ccttgagctt tctcctggca 840

ttgaactttt tcagacactc cacagtctcc tgtctgtgca tcatggatgc taccgtggag 900

cgttggcaga cccacgggtg cttcagggcc tcatgggctg tgatgcgctt ggcagggttg 960

atggtcagca tctggttgat gaggtttttg gcttcaggag tgacggtgtc ccactcaggg 1020

gacgggaagt cataggcacc agccttgatc tgctggtaca gcttgtgctg gtcctcgtcc 1080

cagaagggtg ggtagcccac gagcaggatg tacaggatca ccccacatgc ccagatgtcc 1140

acaggcttgc catacgcctc tttgcgaagg acctcagggg acaggtagcc tggtgtgcca 1200

gcgaaaccaa accatgcctg ctggtccccc tgcacctcga tagctaggcc gaagtctgcc 1260

agcttcactg cagccccttt gcacttgctg gccagaagca ggttctccgg cttgaggtct 1320

ctgtggacga cccccatttg gtgacaatgg agaacggcct ccaggatctg ctggatacag 1380

tgactggcat cagcctcgct gtagtactct ctcgccacaa tgtcttcaaa gagctcccca 1440

ccagtgacca gatcgaagac caggtagtgg aagccctcct cggagatgct gtcgtggaga 1500

cgcacgatgt tggaatgctt cagaaggcgg cagatccgag cctctctctc cagcttctgg 1560

tgatctctgg ctgacagctt cttggtgttg atgatcttgg ctgcatactc atggccggtg 1620

cagagcttga cacagcgtcg gaccacagag aaagcccccc tgccaatatc ctcgtagagc 1680

tggtactcgt cggtgaagcg ggtgcaggtc accgtggtgg ccatggc 1727

<210> 12
<211> 542
<212> PRT
<213> Homo sapiens

<400> 12

Met Ala Thr Thr Val Thr Cys Thr Arg Phe Thr Asp Glu Tyr Gln Leu
1 5 10 15

Tyr Glu Asp Ile Gly Lys Gly Ala Phe Ser Val Val Arg Arg Cys Val
20 25 30

Lys Leu Cys Thr Gly His Glu Tyr Ala Ala Lys Ile Ile Asn Thr Lys
35 40 45

Lys Leu Ser Ala Arg Asp His Gln Lys Leu Glu Arg Glu Ala Arg Ile
50 55 60

Cys Arg Leu Leu Lys His Ser Asn Ile Val Arg Leu His Asp Ser Ile
65 70 75 80

Ser Glu Glu Gly Phe His Tyr Leu Val Phe Asp Leu Val Thr Gly Gly
85 90 95

Glu Leu Phe Glu Asp Ile Val Ala Arg Glu Tyr Tyr Ser Glu Ala Asp
100 105 110

Ala Ser His Cys Ile Gln Gln Ile Leu Glu Ala Val Leu His Cys His
115 120 125

Gln Met Gly Val Val His Arg Asp Leu Lys Pro Glu Asn Leu Leu Leu
130 135 140

Ala Ser Lys Cys Lys Gly Ala Ala Val Lys Leu Ala Asp Phe Gly Leu
145 150 155 160

Ala Ile Glu Val Gln Gly Asp Gln Gln Ala Trp Phe Gly Phe Ala Gly
165 170 175

Thr Pro Gly Tyr Leu Ser Pro Glu Val Leu Arg Lys Glu Ala Tyr Gly
180 185 190

Lys Pro Val Asp Ile Trp Ala Cys Gly Val Ile Leu Tyr Ile Leu Leu
195 200 205

Val Gly Tyr Pro Pro Phe Trp Asp Glu Asp Gln His Lys Leu Tyr Gln
210 215 220

Gln Ile Lys Ala Gly Ala Tyr Asp Phe Pro Ser Pro Glu Trp Asp Thr
225 230 235 240

Val Thr Pro Glu Ala Lys Asn Leu Ile Asn Gln Met Leu Thr Ile Asn
245 250 255

Pro Ala Lys Arg Ile Thr Ala His Glu Ala Leu Lys His Pro Trp Val
260 265 270

Cys Gln Arg Ser Thr Val Ala Ser Met Met His Arg Gln Glu Thr Val
275 280 285

Glu Cys Leu Lys Lys Phe Asn Ala Arg Arg Lys Leu Lys Gly Ala Ile
290 295 300

Leu Thr Thr Met Leu Ala Thr Arg Asn Phe Ser Val Gly Arg Gln Thr
305 310 315 320

Thr Ala Pro Ala Thr Met Ser Thr Ala Ala Ser Gly Thr Thr Met Gly
325 330 335

Leu Val Glu Gln Ala Lys Ser Leu Leu Asn Lys Lys Ala Asp Gly Val
340 345 350

Lys Pro Gln Thr Asn Ser Thr Lys Asn Ser Ala Ala Ala Thr Ser Pro
355 360 365

Lys Gly Thr Leu Pro Pro Ala Ala Leu Glu Pro Gln Thr Thr Val Ile
370 375 380

His Asn Pro Val Asp Gly Ile Lys Glu Ser Ser Asp Ser Ala Asn Thr
385 390 395 400

Thr Ile Glu Asp Glu Asp Ala Lys Ala Arg Lys Gln Glu Ile Ile Lys
405 410 415

Thr Thr Glu Gln Leu Ile Glu Ala Val Asn Asn Gly Asp Phe Glu Ala
420 425 430

Tyr Ala Lys Ile Cys Asp Pro Gly Leu Thr Ser Phe Glu Pro Glu Ala
435 440 445

Leu Gly Asn Leu Val Glu Gly Met Asp Phe His Arg Phe Tyr Phe Glu
450 455 460

Asn Leu Leu Ala Lys Asn Ser Lys Pro Ile His Thr Thr Ile Leu Asn
465 470 475 480

Pro His Val His Val Ile Gly Glu Asp Ala Ala Cys Ile Ala Tyr Ile
485 490 495

Arg Leu Thr Gln Tyr Ile Asp Gly Gln Gly Arg Pro Arg Thr Ser Gln
500 505 510

Ser Glu Glu Thr Arg Val Trp His Arg Arg Asp Gly Lys Trp Gln Asn
515 520 525

Val His Phe His Cys Ser Gly Ala Pro Val Ala Pro Leu Gln
530 535 540

<210> 13
<211> 4114
<212> DNA
<213> Homo sapiens

<400> 13
tgagagtgga ggggattcaa gggggatgac tccagcagag cacctcactc ctttgaagag 60

cacagaggaa gatgtcagcc cagtcccttc ctgcagcaac acccccacg cagaagcccc 120

ctcggatcat ccgcccccgc cctccttctc gttccagggc tgcccagtcc ccagggcctc 180

cccacaatgg ctcctctcca caagaactac cccgaaactc caatgatgca ccaaccccaa 240

tgtgcacccc catcttctgg gagcccccag ctgcatccct caagcccccc gctcttttgc 300

ccccctcagc ttctagagcc agcctcgact cccagacttc cccagactca ccttccagca 360

cccccacacc tagtccagtg tcccggcgct ccgcctcccc agaacctgct ccccggtctc 420

cagtcccccc acccaagccg tctgggtcac cctgcacgcc tctgctcccc atggctggag 480

tcctggctca gaatggctct gcctcagctc ctggcactgt gcggaggctg gctggcaggt 540

ttgaaggggg tgctgaaggc cgggctcagg atgcagatgc cccggagcca ggtctccaag 600

cgagagcaga tgtgaatggg gagagagaag ctcccctcac cgggagtggg tcccaggaga 660

acggtgctcc agatgctggc ctggcctgcc ctccctgctg cccctgtgtc tgccacacca 720

cccggcctgg cctggagctc agatgggtgc ctgtgggggg ctatgaggag gtccccaggg 780

tcccccgtcg ggcctccccg ctgcggacct ctcgctcccg cccccaccct ccaagcatcg 840

gtcaccctgc cgttgtcctc acatcctacc gctccactgc tgagcgcaaa ctcctgccac 900

tcctcaagcc tcccaaacca actcgtgtca ggcaggatgc caccattttc ggggaccccc 960

cacagccaga tcttgatctg ctttctgaag atggaatcca aacaggggac agtcctgatg 1020

aagctcctca gaatactcct ccagcaactg tggaggggag ggaagaggag gggctagagg 1080

tgctgaagga gcagaattgg gagctgcccc tgcaggatga acctctgtac cagacctacc 1140

gagcagccgt gctgtcagag gagctgtggg gggtgggtga ggatgggagt ccttctccag 1200

caaatgctgg agatgcaccc accttcccac gacccccctgg acctcgcaac accctgtggc 1260

aggagcttcc ggctgtgcaa gccagcggtc ttctggatac cctcagcccc caggagaggc 1320

gcatgcagga gagtcttttc gaggtggtga cgtccgaggc ttcctacctg cgctccctgc 1380

ggctgctgac cgacaccttc gtgctgagcc aggcactccg ggacacgctc acccccgtg 1440

atcaccacac actcttctcc aatgtgcagc gagtccaggg agtcagcgag cggtttctag 1500

caacgctcct gtcccgtgtg cgctcttccc cccacatcag cgacttgtgt gatgtggtgc 1560

atgcccacgc tgtggggcct ttctcggtgt atgtggatta tgtgcggaac cagcagtatc 1620

aggaggagac ctacagccgc ctcatggaca ccaacgtgcg cttctccgcc gagctgcgcc 1680

ggctgcagag cctccctaag tgtgagcggc tcccgctgcc gtccttcctg ctactgccct 1740

tccagcgcat cacccggctg cgcatgctgc tgcagaatat cctgcgccag acagaagagg 1800

ggtccagccg tcaggagaat gcccagaagg ccctgggtgc tgtcagcaag atcatcgagc 1860

gttgcagcgc tgaggtgggg cgcatgaagc agactgaaga gctgatccgg ctcacccaaa 1920

ggctgcgctt ccacaaagtc aaggccctgc ccctggtctc ctggtcacgg cgcctggaat 1980

tccagggaga gctgactgag ttagggtgcc ggaggggggg cgtgctcttt gcctcgcgcc 2040

cccgcttcac ccctctttgc ctgctgctct ttagcgacct gctgctcatc actcagccta 2100

agagtgggca gcggttacag gttctggact atgcccatcg ctccctggtc caggcccagc 2160

aggttccgga tccatctgga ccccctacct tccgcctctc ccttctcagc aaccaccagg 2220

gccgccccac ccaccgacta ctccaagctt cttccctatc agacatgcag cgctggctgg 2280

gagccttccc aaccccaggc ccccttccct gctccccaga caccatctat gaggactgtg 2340

actgttccca ggaactgtgt tcagagtcgt ctgcacctgc caagactgaa ggacggagtc 2400

tggagtccag ggctgccccc aaacacctgc acaagacccc tgaaggttgg ctgaaggggc 2460

ttcctggggc cttccctgcc cagctggtgt gtgaagtcac aggggaacac gaaaggagga 2520

ggcaccttcg ccagaaccag aggcttctcg aggctgttgg accttcttca ggcaccccca 2580

atgcccccc accctaatgc aggctgagga gggggcacat gttgggagac acctaccagt 2640

gtggcacgga gagaacaaag cccattcatc cattggattc actgtcagtg gagatactac 2700

ctctcgtggc aaccatagag atcgagcttc aggacagagc agccaatgaa aacggccgcc 2760

tgaacccaca gcaataagaa tgaatgagga tgccttgaat gtgtggccaa tggagacaga 2820

ggcttagtgc agagcagcca atgggtactg agctggctga gcctatggcc aatgagtatt 2880

cctgctatgc tcagggccaa ggaagacaaa tctaggtcat ggcagttgaa aaagggcctc 2940

attggagata aagtcgtagg ataaaattgg gaacaggaat gagcaggaag ccaatcagcc 3000

aaagaaaatg gtgatctgga cccaagagac cagtagtcac cctgcttgtt tctgcagcaa 3060

tgactggtcc tgtcttttga gtctgggaaa tactagtttc cattcctgga tgcttcctgt 3120

gccctctcaa gccagttctt ctcttccaga agaattcaga gtatgtgtct cagaaaatct 3180

gtgtgtgtac atgtgcatgt gtagatatgt gtgtatatgt atcaggaaag gcattctgct 3240

gactgtggtg tgtgtgtggt gattgtgctc ctgacccaca aatgactgag tgctccattt 3300

cttcctttac cccccatttt tcctattatt cgctccaaga aagatgctaa gtctgagctc 3360

cagaagagac tgtgctgggt gtggcttggc acccagggga tgagagccct gagctttggg 3420

tctcttggag gctagggttc tgtggcagtt gcagggcaat gttatggagc agccaacggc 3480

ctggcagagg agcccaaggg actgaagatg gccagtagct gggtcctgag gcccctgaag 3540

tctgcagacc cttctccttg ccccaaacac tggcctccat aattcctgcc tgcagatctc 3600

ccaacttgaa ctataatcca ccagccagcc tcagccttga gctttggaac cacattagat 3660

cctgcatctg ggtgaagaaa cgggagctgt ggaccacagg ccagccagtg aacctcctgg 3720

gctttcttgc ctttgtcctg atcctctcac agaaacactg ggccaaacag tggggagaga 3780

ttggagagcg ggtgtggctg cccaacccca tccagagcat ctgcttccag atgagccagt 3840

gcctcgcatg ataccagagg aggcgaggga cagagacagc aaggcagaca gtggctggca 3900

gggggggccca ggcccgggac gaggcctccc cttcagctca ggcacagcaa cttgcccagg 3960

actgacactg tcaccctgac tgcaggaggc acagggactc cgggagactc agagggcgaa 4020

gagcactggc atttggcatg tccatgacat tggagactcc cctagcaggg tgcctgacgt 4080

gtggggaacc ctcagtaaat agtggtgcat ttgt 4114

<210> 14
<211> 4114
<212> DNA
<213> Homo sapiens

<400> 14
acaaatgcac cactatttac tgagggttcc ccacacgtca ggcaccctgc taggggagtc 60

tccaatgtca tggacatgcc aaatgccagt gctcttcgcc ctctgagtct cccggagtcc 120

ctgtgcctcc tgcagtcagg gtgacagtgt cagtcctggg caagttgctg tgcctgagct 180

gaaggggagg cctcgtcccg ggcctgggcc cccctgccag ccactgtctg ccttgctgtc 240

tctgtccctc gcctcctctg gtatcatgcg aggcactggc tcatctggaa gcagatgctc 300

tggatggggt tgggcagcca cacccgctct ccaatctctc cccactgttt ggcccagtgt 360

ttctgtgaga ggatcaggac aaaggcaaga aagcccagga ggttcactgg ctggcctgtg 420

gtccacagct cccgtttctt cacccagatg caggatctaa tgtggttcca aagctcaagg 480

ctgaggctgg ctggtggatt atagttcaag ttgggagatc tgcaggcagg aattatggag 540

gccagtgttt ggggcaagga gaagggtctg cagacttcag gggcctcagg acccagctac 600

tggccatctt cagtcccttg ggctcctctg ccaggccgtt ggctgctcca taacattgcc 660

ctgcaactgc cacagaaccc tagcctccaa gagacccaaa gctcagggct ctcatcccct 720

gggtgccaag ccacacccag cacagtctct tctggagctc agacttagca tctttcttgg 780

agcgaataat aggaaaaatg gggggtaaag gaagaaatgg agcactcagt catttgtggg 840

tcaggagcac aatcaccaca cacacaccac agtcagcaga atgcctttcc tgatacatat 900

acacacatat ctacacatgc acatgtacac acacagattt tctgagacac atactctgaa 960

ttcttctgga agagaagaac tggcttgaga gggcacagga agcatccagg aatggaaact 1020

agtatttccc agactcaaaa gacaggacca gtcattgctg cagaaacaag cagggtgact 1080

actggtctct tgggtccaga tcaccatttt ctttggctga ttggcttcct gctcattcct 1140

gttcccaatt ttatcctacg actttatctc caatgaggcc ctttttcaac tgccatgacc 1200

tagatttgtc ttccttggcc ctgagcatag caggaatact cattggccat aggctcagcc 1260

agctcagtac ccattggctg ctctgcacta agcctctgtc tccattggcc acacattcaa 1320

ggcatcctca ttcattctta ttgctgtggg ttcaggcggc cgttttcatt ggctgctctg 1380

tcctgaagct cgatctctat ggttgccacg agaggtagta tctccactga cagtgaatcc 1440

aatggatgaa tgggctttgt tctctccgtg ccacactggt aggtgtctcc caacatgtgc 1500

cccctcctca gcctgcatta gggtgggggg gcattggggg tgcctgaaga aggtccaaca 1560

gcctcgagaa gcctctggtt ctggcgaagg tgcctcctcc tttcgtgttc ccctgtgact 1620

tcacacacca gctgggcagg gaaggcccca ggaagcccct tcagccaacc ttcaggggtc 1680

ttgtgcaggt gtttgggggc agccctggac tccagactcc gtccttcagt cttggcaggt 1740

gcagacgact ctgaacacag ttcctgggaa cagtcacagt cctcatagat ggtgtctggg 1800

gagcagggaa gggggcctgg ggttgggaag gctcccagcc agcgctgcat gtctgatagg 1860

gaagaagctt ggagtagtcg gtgggtgggg cggccctggt ggttgctgag aagggagagg 1920

cggaaggtag ggggtccaga tggatccgga acctgctggg cctggaccag ggagcgatgg 1980

gcatagtcca gaacctgtaa ccgctgccca ctcttaggct gagtgatgag cagcaggtcg 2040

ctaaagagca gcaggcaaag aggggtgaag cgggggcgcg aggcaaagag cacgcccccc 2100

ctccggcacc ctaactcagt cagctctccc tggaattcca ggcgccgtga ccaggagacc 2160

aggggcaggg ccttgacttt gtggaagcgc agcctttggg tgagccggat cagctcttca 2220

gtctgcttca tgcgccccac ctcagcgctg caacgctcga tgatcttgct gacagcaccc 2280

agggccttct gggcattctc ctgacggctg gacccctctt ctgtctggcg caggatattc 2340

tgcagcagca tgcgcagccg ggtgatgcgc tggaagggca gtagcaggaa ggacggcagc 2400

gggagccgct cacacttagg gaggctctgc agccggcgca gctcggcgga gaagcgcacg 2460

ttggtgtcca tgaggcggct gtaggtctcc tcctgatact gctggttccg cacataatcc 2520

acatacaccg agaaaggccc cacagcgtgg gcatgcacca catcacacaa gtcgctgatg 2580

tgggggggaag agcgcacacg ggacaggagc gttgctagaa accgctcgct gactccctgg 2640

actcgctgca cattggagaa gagtgtgtgg tgatcacggg gggtgagcgt gtcccggagt 2700

gcctggctca gcacgaaggt gtcggtcagc agccgcaggg agcgcaggta ggaagcctcg 2760

gacgtcacca cctcgaaaag actctcctgc atgcgcctct cctgggggct gagggtatcc 2820

agaagaccgc tggcttgcac agccggaagc tcctgccaca gggtgttgcg aggtccaggg 2880

ggtcgtggga aggtgggtgc atctccagca tttgctggag aaggactccc atcctcaccc 2940

acccccaca gctcctctga cagcacggct gctcggtagg tctggtacag aggttcatcc 3000

tgcaggggca gctcccaatt ctgctccttc agcacctcta gcccctcctc ttccctcccc 3060

tccacagttg ctggaggagt attctgagga gcttcatcag gactgtcccc tgtttggatt 3120

ccatcttcag aaagcagatc aagatctggc tgtgggggggt ccccgaaaat ggtggcatcc 3180

tgcctgacac gagttggttt gggaggcttg aggagtggca ggagtttgcg ctcagcagtg 3240

gagcggtagg atgtgaggac aacggcaggg tgaccgatgc ttggagggtg ggggcgggag 3300

cgagaggtcc gcagcgggga ggcccgacgg gggaccctgg ggacctcctc atagcccccc 3360

acaggcaccc atctgagctc caggccaggc cgggtggtgt ggcagacaca ggggcagcag 3420

ggagggcagg ccaggccagc atctggagca ccgttctcct gggacccact cccggtgagg 3480

ggagcttctc tctccccatt cacatctgct ctcgcttgga gacctggctc cggggcatct 3540

gcatcctgag cccggccttc agcacccccct tcaaacctgc cagccagcct ccgcacagtg 3600

ccaggagctg aggcagagcc attctgagcc aggactccag ccatggggag cagaggcgtg 3660

cagggtgacc cagacggctt gggtgggggg actggagacc ggggagcagg ttctggggag 3720

gcggagcgcc gggacactgg actaggtgtg ggggtgctgg aaggtgagtc tggggaagtc 3780

tgggagtcga ggctggctct agaagctgag gggggcaaaa gagcaggggg cttgagggat 3840

gcagctgggg gctcccagaa gatgggggtg cacattgggg ttggtgcatc attggagttt 3900

cggggtagtt cttgtggaga ggagccattg tggggaggcc ctggggactg ggcagccctg 3960

gaacgagaag gagggcgggg gcggatgatc cgagggggct tctgcgtggg gggtgttgct 4020

gcaggaaggg actgggctga catcttcctc tgtgctcttc aaaggagtga ggtgctctgc 4080

tggagtcatc ccccttgaat cccctccact ctca 4114

<210> 15
<211> 841
<212> PRT
<213> Homo sapiens

<400> 15

```
Met Ser Ala Gln Ser Leu Pro Ala Ala Thr Pro Pro Thr Gln Lys Pro
1               5                  10                  15

Pro Arg Ile Ile Arg Pro Arg Pro Pro Ser Arg Ser Arg Ala Ala Gln
            20                  25                  30

Ser Pro Gly Pro Pro His Asn Gly Ser Ser Pro Gln Glu Leu Pro Arg
        35                  40                  45

Asn Ser Asn Asp Ala Pro Thr Pro Met Cys Thr Pro Ile Phe Trp Glu
    50                  55                  60

Pro Pro Ala Ala Ser Leu Lys Pro Pro Ala Leu Leu Pro Pro Ser Ala
65                  70                  75                  80

Ser Arg Ala Ser Leu Asp Ser Gln Thr Ser Pro Asp Ser Pro Ser Ser
                85                  90                  95

Thr Pro Thr Pro Ser Pro Val Ser Arg Arg Ser Ala Ser Pro Glu Pro
            100                 105                 110

Ala Pro Arg Ser Pro Val Pro Pro Pro Lys Pro Ser Gly Ser Pro Cys
        115                 120                 125

Thr Pro Leu Leu Pro Met Ala Gly Val Leu Ala Gln Asn Gly Ser Ala
    130                 135                 140

Ser Ala Pro Gly Thr Val Arg Arg Leu Ala Gly Arg Phe Glu Gly Gly
145                 150                 155                 160

Ala Glu Gly Arg Ala Gln Asp Ala Asp Ala Pro Glu Pro Gly Leu Gln
                165                 170                 175

Ala Arg Ala Asp Val Asn Gly Glu Arg Glu Ala Pro Leu Thr Gly Ser
            180                 185                 190

Gly Ser Gln Glu Asn Gly Ala Pro Asp Ala Gly Leu Ala Cys Pro Pro
        195                 200                 205

Cys Cys Pro Cys Val Cys His Thr Thr Arg Pro Gly Leu Glu Leu Arg
    210                 215                 220

Trp Val Pro Val Gly Gly Tyr Glu Glu Val Pro Arg Val Pro Arg Arg
225                 230                 235                 240
```

Ala Ser Pro Leu Arg Thr Ser Arg Ser Arg Pro His Pro Pro Ser Ile
245 250 255

Gly His Pro Ala Val Val Leu Thr Ser Tyr Arg Ser Thr Ala Glu Arg
260 265 270

Lys Leu Leu Pro Leu Leu Lys Pro Pro Lys Pro Thr Arg Val Arg Gln
275 280 285

Asp Ala Thr Ile Phe Gly Asp Pro Pro Gln Pro Asp Leu Asp Leu Leu
290 295 300

Ser Glu Asp Gly Ile Gln Thr Gly Asp Ser Pro Asp Glu Ala Pro Gln
305 310 315 320

Asn Thr Pro Pro Ala Thr Val Glu Gly Arg Glu Glu Glu Gly Leu Glu
325 330 335

Val Leu Lys Glu Gln Asn Trp Glu Leu Pro Leu Gln Asp Glu Pro Leu
340 345 350

Tyr Gln Thr Tyr Arg Ala Ala Val Leu Ser Glu Glu Leu Trp Gly Val
355 360 365

Gly Glu Asp Gly Ser Pro Ser Pro Ala Asn Ala Gly Asp Ala Pro Thr
370 375 380

Phe Pro Arg Pro Pro Gly Pro Arg Asn Thr Leu Trp Gln Glu Leu Pro
385 390 395 400

Ala Val Gln Ala Ser Gly Leu Leu Asp Thr Leu Ser Pro Gln Glu Arg
405 410 415

Arg Met Gln Glu Ser Leu Phe Glu Val Val Thr Ser Glu Ala Ser Tyr
420 425 430

Leu Arg Ser Leu Arg Leu Leu Thr Asp Thr Phe Val Leu Ser Gln Ala
435 440 445

Leu Arg Asp Thr Leu Thr Pro Arg Asp His His Thr Leu Phe Ser Asn
450 455 460

Val Gln Arg Val Gln Gly Val Ser Glu Arg Phe Leu Ala Thr Leu Leu
465 470 475 480

Ser Arg Val Arg Ser Ser Pro His Ile Ser Asp Leu Cys Asp Val Val
485 490 495

His Ala His Ala Val Gly Pro Phe Ser Val Tyr Val Asp Tyr Val Arg
500 505 510

Asn Gln Gln Tyr Gln Glu Glu Thr Tyr Ser Arg Leu Met Asp Thr Asn
515 520 525

Val Arg Phe Ser Ala Glu Leu Arg Arg Leu Gln Ser Leu Pro Lys Cys
530 535 540

Glu Arg Leu Pro Leu Pro Ser Phe Leu Leu Leu Pro Phe Gln Arg Ile
545 550 555 560

Thr Arg Leu Arg Met Leu Leu Gln Asn Ile Leu Arg Gln Thr Glu Glu
565 570 575

Gly Ser Ser Arg Gln Glu Asn Ala Gln Lys Ala Leu Gly Ala Val Ser
580 585 590

Lys Ile Ile Glu Arg Cys Ser Ala Glu Val Gly Arg Met Lys Gln Thr
595 600 605

Glu Glu Leu Ile Arg Leu Thr Gln Arg Leu Arg Phe His Lys Val Lys
610 615 620

Ala Leu Pro Leu Val Ser Trp Ser Arg Arg Leu Glu Phe Gln Gly Glu
625 630 635 640

Leu Thr Glu Leu Gly Cys Arg Arg Gly Gly Val Leu Phe Ala Ser Arg
645 650 655

Pro Arg Phe Thr Pro Leu Cys Leu Leu Leu Phe Ser Asp Leu Leu Leu
660 665 670

Ile Thr Gln Pro Lys Ser Gly Gln Arg Leu Gln Val Leu Asp Tyr Ala
675 680 685

His Arg Ser Leu Val Gln Ala Gln Gln Val Pro Asp Pro Ser Gly Pro
690 695 700

Pro Thr Phe Arg Leu Ser Leu Leu Ser Asn His Gln Gly Arg Pro Thr
705 710 715 720

His Arg Leu Leu Gln Ala Ser Ser Leu Ser Asp Met Gln Arg Trp Leu
725 730 735

Gly Ala Phe Pro Thr Pro Gly Pro Leu Pro Cys Ser Pro Asp Thr Ile
740 745 750

Tyr Glu Asp Cys Asp Cys Ser Gln Glu Leu Cys Ser Glu Ser Ser Ala
755 760 765

Pro Ala Lys Thr Glu Gly Arg Ser Leu Glu Ser Arg Ala Ala Pro Lys
770 775 780

His Leu His Lys Thr Pro Glu Gly Trp Leu Lys Gly Leu Pro Gly Ala
785 790 795 800

Phe Pro Ala Gln Leu Val Cys Glu Val Thr Gly Glu His Glu Arg Arg
805 810 815

Arg His Leu Arg Gln Asn Gln Arg Leu Leu Glu Ala Val Gly Pro Ser
820 825 830

Ser Gly Thr Pro Asn Ala Pro Pro Pro
835 840

<210> 16
<211> 2355
<212> DNA
<213> Homo sapiens

<400> 16
agagacagcc cagcctgggc catggaagaa aacccgacct tggaatcaga agcctggggc 60

tcctctaggg ggtggctggc cccccgggag gccagaggag gcccatcgct gtcttctgtg 120

ctgaacgagc tgcccagtgc tgccaccctt cggtaccgag accctggggt gctgccttgg 180

ggggcgctgg aggaggagga ggaggatgga ggaaggagca gaaaggcctt cacagaagtc 240

acccagacag agctgcagga ccctcaccct tcccgggaac tgccctggcc catgcaggcc 300

agacgggcac acaggcaaag aaatgccagc agggaccagg tggtctatgg ctctggaact 360

aagacggacc gatgggcgcg gctacttcgg aggtccaagg agaaaacaaa ggaaggcttg 420

cgaagcctgc agccctgggc gtggacactg aagaggatcg ggggccagtt tggcgccggc 480

acggagtcct acttctccct gctgcgcttc ctgctccttc ttaacgtgct ggcctctgtg 540

ctcatggcct gcatgacgct gctgcccacc tggttgggag gcgctccccc aggccctccc 600

ggccccgaca tctcctcgcc ctgcggctcc tataaccccc actcccaggg cctggtcacc 660

tttgccaccc agctcttcaa cttgctctcg ggtgagggtt acctggaatg gtcccctctc 720

ttctatggct tctacccgcc ccgcccacgc ctggcggtca cctacctgtg ctgggccttt 780

gccgttggcc tcatctgcct cctgctcatc ctgcatcgct cggtgtctgg gctgaagcag 840

acactgctgg cggagtccga ggctctgacc agctacagcc accgggtgtt ctcggcctgg 900

gacttcggtc tctgcgggga cgtccacgtg cggctgcgcc agcgcatcat cttgtacgaa 960

ttaaaggtgg agctggagga gacagtggtg cggcgccagg ctgcggtgcg gacgctgggc 1020

cagcaagcca gggtttggtt ggtgcgggtg ctgctcaacc tgctggtggt cgcgctcctg 1080

ggggcagcct tctatggcgt ctactgggct acggggtgca ccgtggagct gcaggagatg 1140

ccccttgtcc aggagttgcc actgctgaag cttggggtga attaccttcc gtccatcttc 1200

atcgctgggg tcaattttgt gctgccgccc gtgttcaagc tcattgctcc actggagggc 1260

tacactcgga gtcgccagat cgtttttatc ctgctcagga ccgtgtttct tcgcctcgcc 1320

tccctggtgg tcctgctctt ctctctctgg aatcagatca cttgtggggg cgactccgag 1380

gctgaggact gcaaaacctg tggctacaat tacaaacaac ttccgtgctg ggagactgtc 1440

ctgggccagg aaatgtacaa acttctgctc tttgatctgc tgactgtctt ggcagtcgcg 1500

ctgctcatcc agtttcctag aaagctcctc tgtggcctct gtcctggggc gctgggtctt 1560

ctggcgggga cccaggagtt ccaggtgccc gacgaggtgc tggggctcat ctacgcgcag 1620

acggtggtct gggtggggag ttttttctgc cctttactgc ccctgcttaa cacggtcaag 1680

ttcctgctgc ttttctacct gaagaagctt accctcttct ccacctgctc cccggctgcc 1740

cgcaccttcc gggcctccgc ggcgaatttc tttttccccc tggtccttct cctgggtctg 1800

gccatctcca gcgttccccc gctttacagc atcttcctga tcccgccttc taagctgtgt 1860

ggtccattcc gggggcagtc gtccatctgg gcccagatcc ctgagtctat ttccagcctc 1920

cctgagacca cccagaattt cctcttcttc ctggggaccc aggcttttgc tgtgcccctt 1980

ctgctgatct ccagcatcct gatggcgtac actgtggctc tggctaactc ctacggacgc 2040

ctcatctctg agctcaaacg tcagagagag acggaggcgc agaataaagt cttcctggca 2100

cggcgcgctg tggcgctgac ctccaccaaa ccggctcttt gacccccgca gcccacgtcc 2160

cgctttcaga ccccaggccc attgtaagcc taggtcacaa catctgtaaa ctaggagaac 2220

tggagaagac tccacgccct tccagctttg gtatctggag atttccaggg cccctcgccg 2280

ccacgtccct gactctcggg tgatcttcct tgtatcaata aatacagccg aggttgcaaa 2340

aaaaaaaaaa aaaaa 2355

<210> 17
<211> 2355
<212> DNA
<213> Homo sapiens

<400> 17
ttttttttttt ttttttttgc aacctcggct gtatttattg atacaaggaa gatcacccga 60

gagtcaggga cgtggcggcg aggggccctg gaaatctcca gataccaaag ctggaagggc 120

gtggagtctt ctccagttct cctagtttac agatgttgtg acctaggctt acaatgggcc 180

tggggtctga aagcgggacg tgggctgcgg gggtcaaaga gccggtttgg tggaggtcag 240

cgccacagcg cgccgtgcca ggaagacttt attctgcgcc tccgtctctc tctgacgttt 300
gagctcagag atgaggcgtc cgtaggagtt agccagagcc acagtgtacg ccatcaggat 360
gctggagatc agcagaaggg gcacagcaaa agcctgggtc cccaggaaga agaggaaatt 420
ctgggtggtc tcagggaggc tggaaataga ctcagggatc tgggcccaga tggacgactg 480
cccccggaat ggaccacaca gcttagaagg cgggatcagg aagatgctgt aaagcagggg 540
aacgctggag atggccagac ccaggagaag gaccaagggg aaaaagaaat tcgccgcgga 600
ggcccggaag gtgcgggcag ccggggagca ggtggagaag agggtaagct tcttcaggta 660
gaaaagcagc aggaacttga ccgtgttaag caggggcagt aaagggcaga aaaaactccc 720
cacccagacc accgtctgcg cgtagatgag ccccagcacc tcgtcgggca cctggaactc 780
ctgggtcccc gccagaagac ccagcgcccc aggacagagg ccacagagga gctttctagg 840
aaactggatg agcagcgcga ctgccaagac agtcagcaga tcaaagagca gaagtttgta 900
catttcctgg cccaggacag tctcccagca cggaagttgt ttgtaattgt agccacaggt 960
tttgcagtcc tcagcctcgg agtcgccccc acaagtgatc tgattccaga gagagaagag 1020
caggaccacc agggaggcga ggcgaagaaa cacggtcctg agcaggataa aaacgatctg 1080
gcgactccga gtgtagccct ccagtggagc aatgagcttg aacacgggcg gcagcacaaa 1140
attgacccca gcgatgaaga tggacggaag gtaattcacc ccaagcttca gcagtggcaa 1200
ctcctggaca aggggcatct cctgcagctc cacggtgcac cccgtagccc agtagacgcc 1260
atagaaggct gcccccagga gcgcgaccac cagcaggttg agcagcaccc gcaccaacca 1320
aaccctggct tgctggccca gcgtccgcac cgcagcctgg cgccgcacca ctgtctcctc 1380
cagctccacc tttaattcgt acaagatgat gcgctggcgc agccgcacgt ggacgtcccc 1440
gcagagaccg aagtcccagg ccgagaacac ccggtggctg tagctggtca gagcctcgga 1500
ctccgccagc agtgtctgct tcagcccaga caccgagcga tgcaggatga gcaggaggca 1560
gatgaggcca acggcaaagg cccagcacag gtaggtgacc gccaggcgtg ggcggggcgg 1620
gtagaagcca tagaagagag gggaccattc caggtaaccc tcacccgaga gcaagttgaa 1680
gagctgggtg gcaaaggtga ccaggccctg ggagtggggg ttataggagc cgcagggcga 1740
ggagatgtcg gggccgggag ggcctgggggg agcgcctccc aaccaggtgg gcagcagcgt 1800
catgcaggcc atgagcacag aggccagcac gttaagaagg agcaggaagc gcagcaggga 1860
gaagtaggac tccgtgccgg cgccaaactg gcccccgatc ctcttcagtg tccacgccca 1920
gggctgcagg cttcgcaagc cttcctttgt tttctccttg gacctccgaa gtagccgcgc 1980
ccatcggtcc gtcttagttc cagagccata gaccacctgg tccctgctgg catttctttg 2040
cctgtgtgcc cgtctggcct gcatgggcca gggcagttcc cgggaagggt gagggtcctg 2100

cagctctgtc tgggtgactt ctgtgaaggc ctttctgctc cttcctccat cctcctcctc 2160

ctcctccagc gccccccaag gcagcacccc agggtctcgg taccgaaggg tggcagcact 2220

gggcagctcg ttcagcacag aagacagcga tgggcctcct ctggcctccc gggggggccag 2280

ccacccccta gaggagcccc aggcttctga ttccaaggtc gggttttctt ccatggccca 2340

ggctgggctg tctct 2355

<210> 18
<211> 713
<212> PRT
<213> Homo sapiens

<400> 18

Arg Asp Ser Pro Ala Trp Ala Met Glu Glu Asn Pro Thr Leu Glu Ser
1 5 10 15

Glu Ala Trp Gly Ser Ser Arg Gly Trp Leu Ala Pro Arg Glu Ala Arg
20 25 30

Gly Gly Pro Ser Leu Ser Ser Val Leu Asn Glu Leu Pro Ser Ala Ala
35 40 45

Thr Leu Arg Tyr Arg Asp Pro Gly Val Leu Pro Trp Gly Ala Leu Glu
50 55 60

Glu Glu Glu Glu Asp Gly Gly Arg Ser Arg Lys Ala Phe Thr Glu Val
65 70 75 80

Thr Gln Thr Glu Leu Gln Asp Pro His Pro Ser Arg Glu Leu Pro Trp
85 90 95

Pro Met Gln Ala Arg Arg Ala His Arg Gln Arg Asn Ala Ser Arg Asp
100 105 110

Gln Val Val Tyr Gly Ser Gly Thr Lys Thr Asp Arg Trp Ala Arg Leu
115 120 125

Leu Arg Arg Ser Lys Glu Lys Thr Lys Glu Gly Leu Arg Ser Leu Gln
130 135 140

Pro Trp Ala Trp Thr Leu Lys Arg Ile Gly Gly Gln Phe Gly Ala Gly
145 150 155 160

Thr Glu Ser Tyr Phe Ser Leu Leu Arg Phe Leu Leu Leu Leu Asn Val
165 170 175

Leu Ala Ser Val Leu Met Ala Cys Met Thr Leu Leu Pro Thr Trp Leu

180 185 190

Gly Gly Ala Pro Pro Gly Pro Pro Gly Pro Asp Ile Ser Ser Pro Cys
195 200 205

Gly Ser Tyr Asn Pro His Ser Gln Gly Leu Val Thr Phe Ala Thr Gln
210 215 220

Leu Phe Asn Leu Leu Ser Gly Glu Gly Tyr Leu Glu Trp Ser Pro Leu
225 230 235 240

Phe Tyr Gly Phe Tyr Pro Pro Arg Pro Arg Leu Ala Val Thr Tyr Leu
245 250 255

Cys Trp Ala Phe Ala Val Gly Leu Ile Cys Leu Leu Leu Ile Leu His
260 265 270

Arg Ser Val Ser Gly Leu Lys Gln Thr Leu Leu Ala Glu Ser Glu Ala
275 280 285

Leu Thr Ser Tyr Ser His Arg Val Phe Ser Ala Trp Asp Phe Gly Leu
290 295 300

Cys Gly Asp Val His Val Arg Leu Arg Gln Arg Ile Ile Leu Tyr Glu
305 310 315 320

Leu Lys Val Glu Leu Glu Glu Thr Val Val Arg Arg Gln Ala Ala Val
325 330 335

Arg Thr Leu Gly Gln Gln Ala Arg Val Trp Leu Val Arg Val Leu Leu
340 345 350

Asn Leu Leu Val Val Ala Leu Leu Gly Ala Ala Phe Tyr Gly Val Tyr
355 360 365

Trp Ala Thr Gly Cys Thr Val Glu Leu Gln Glu Met Pro Leu Val Gln
370 375 380

Glu Leu Pro Leu Leu Lys Leu Gly Val Asn Tyr Leu Pro Ser Ile Phe
385 390 395 400

Ile Ala Gly Val Asn Phe Val Leu Pro Pro Val Phe Lys Leu Ile Ala
405 410 415

Pro Leu Glu Gly Tyr Thr Arg Ser Arg Gln Ile Val Phe Ile Leu Leu
420 425 430

```
Arg Thr Val Phe Leu Arg Leu Ala Ser Leu Val Val Leu Leu Phe Ser
        435                 440                 445

Leu Trp Asn Gln Ile Thr Cys Gly Gly Asp Ser Glu Ala Glu Asp Cys
    450                 455                 460

Lys Thr Cys Gly Tyr Asn Tyr Lys Gln Leu Pro Cys Trp Glu Thr Val
465                 470                 475                 480

Leu Gly Gln Glu Met Tyr Lys Leu Leu Leu Phe Asp Leu Leu Thr Val
                485                 490                 495

Leu Ala Val Ala Leu Leu Ile Gln Phe Pro Arg Lys Leu Leu Cys Gly
            500                 505                 510

Leu Cys Pro Gly Ala Leu Gly Leu Leu Ala Gly Thr Gln Glu Phe Gln
        515                 520                 525

Val Pro Asp Glu Val Leu Gly Leu Ile Tyr Ala Gln Thr Val Val Trp
    530                 535                 540

Val Gly Ser Phe Phe Cys Pro Leu Leu Pro Leu Leu Asn Thr Val Lys
545                 550                 555                 560

Phe Leu Leu Leu Phe Tyr Leu Lys Lys Leu Thr Leu Phe Ser Thr Cys
                565                 570                 575

Ser Pro Ala Ala Arg Thr Phe Arg Ala Ser Ala Ala Asn Phe Phe Phe
            580                 585                 590

Pro Leu Val Leu Leu Leu Gly Leu Ala Ile Ser Ser Val Pro Leu Leu
        595                 600                 605

Tyr Ser Ile Phe Leu Ile Pro Pro Ser Lys Leu Cys Gly Pro Phe Arg
    610                 615                 620

Gly Gln Ser Ser Ile Trp Ala Gln Ile Pro Glu Ser Ile Ser Ser Leu
625                 630                 635                 640

Pro Glu Thr Thr Gln Asn Phe Leu Phe Phe Leu Gly Thr Gln Ala Phe
                645                 650                 655

Ala Val Pro Leu Leu Leu Ile Ser Ser Ile Leu Met Ala Tyr Thr Val
            660                 665                 670

Ala Leu Ala Asn Ser Tyr Gly Arg Leu Ile Ser Glu Leu Lys Arg Gln
        675                 680                 685
```

Arg Glu Thr Glu Ala Gln Asn Lys Val Phe Leu Ala Arg Arg Ala Val
690 695 700

Ala Leu Thr Ser Thr Lys Pro Ala Leu
705 710

<210> 19
<211> 4125
<212> DNA
<213> Homo sapiens

<400> 19
ttttttttc tattattctt ttactatttt ttctattacc atttttttcta gtaccatttt 60

ttctattatt cttttactat aattgtatat aatatggcag ctgcttgcca catgtactat 120

gtggagagat gtaccaccct gcatcagctt ttaccctaca gaaggaaatc agcgttccat 180

tatattttat tgttatcaac agtttaggaa tacatagctt tgcttttgcc tttttctttc 240

cttccccttg tttcccctcg cctcagagaa aagaaggaaa aaaaaattca tctttcctac 300

ccccctcttt ttggatgata ggacttgaag acaatctgaa ataccacata aactcacttc 360

cagatgtttt ttgtttcata tgcaattgaa ttgggctcag actgtgtttt taagctgtat 420

ggtaaaaata tcactgtctt ctagggcctt attgggggggc agggagagac gtgacacttt 480

gtcagaaggg attgagtctg ctaacttaaa ctttccttga ttcaggaata caaagtctcc 540

agctgtgaac agagactcat cagtgaaata gagtacaggc tagaaaggtc tcctgtggat 600

gaatcaggtg atgaattcac gtatggagat gtgcctgtgg aaaacggaat ggcaccattc 660

tttgagatga agctgaaaca ttacaagatc tttgagggaa tgccagtaac tttcacatgt 720

agagtggctg gaaatccaaa gccaaagatc tattggttta aagatgggaa gcagatctct 780

ccaaagagtg atcactacac cattcaaaga gatctcgatg ggacctgctc cctccatacc 840

acagcctcca ccctagatga tgatgggaat tatacaatta tggctgcaaa ccctcagggc 900

cgcatcagtt gtactggacg gctaatggta caggctgtca accaaagagg tcgaagtccc 960

cggtctccct caggccatcc tcatgtcaga aggcctcgtt ctagatcaag ggacagtgga 1020

gacgaaaatg aaccaattca ggagcgattc ttcagacctc acttcttgca ggctcctgga 1080

gatctgactg ttcaagaagg aaaactctgc agaatggact gcaaagtcag tgggttacca 1140

accccagatc taagctggca actagatgga aagcccgtac gccctgacag tgctcacaag 1200

atgctggtgc gtgagaacgg ggtgcactct ctgatcatag agccagtcac gtcacgtgat 1260

gccggcatct acacatgtat agctaccaac cgagcaggac agaactcatt cagcctggag 1320

cttgtggttg ctgctaaaga agcacacaaa ccccctgtgt ttattgagaa gctccaaaac 1380

acaggagttg ctgatgggta cccagtgcgg ctggaatgtc gtgtattggg agtgccacca 1440

cctcagatat tttggaagaa agaaaatgaa tcactcactc acagcactga ccgagtgagc 1500

atgcaccagg acaaccacgg ctacatctgc ctgctcattc agggagccac aaaagaagat 1560

gctgggtggt atactgtgtc agccaagaat gaagcaggga ttgtgtcctg tactgccagg 1620

ctggacgttt acacccagtg gcatcagcag tcacagagca ccaagccaaa aaaagtacgg 1680

ccctcagcca gtcgctatgc agcactttcg gaccagggac tagacatcaa agcagcgttc 1740

caacctgagg ccaacccatc tcacctgaca ctgaatactg ccttggtaga aagtgaggac 1800

ctgtaatcca gcattcttgt taaagctgaa acactgaaac agccattgcc ttgaccaaca 1860

tattcctttg tcacattatg taaaaggcag aaacatacct ttgactataa gaaattaaaa 1920

aaaaacacca aaataatatt tttcttactt gatataccaa acttagttta agtagataat 1980

gctaatacaa atatacacat tgcacagaaa atacacattt actgtccaat ttaaaacttt 2040

ggaattgctg tgattaaagt gatcaaaatg ccaaaatact aaaggaaatc aattgttcac 2100

aggtaactac aatttgtatt atctacaagt gcctttaaac acaagatata ggtgctgtgt 2160

agcctgatag tgtgaaatgt ttaatgaggg agttgtacca caaacagtac tacaatgatt 2220

ctgaagcaca gtgtattcag acagatacag tgaaccaagt gcaatatgta aggatgaaag 2280

aagaagagat gacaaagaaa tccaagtaaa tgccttgtct ttgcaaatgt ttttatatta 2340

aatcataagg aaggaactac ttgccttaaa tgttaatatc aaaagagttt tctaacaagg 2400

ttaatacctt agttcttaac attttttttc tttatgtgta gtgttttcat gctaccttgg 2460

taggaaactt atttacaaac catattaaaa ggctaattta aatataaata atataaagtg 2520

ctctgaataa agcagaaata tattacagtt cattccacag aaagcatcca aaccacccaa 2580

atgaccaagg catatatagt atttggagga atcaggggtt tggaaggagt agggaggaga 2640

atgaaggaaa atgcaaccag catgattata gtgtgttcat ttagataaaa gtagaaggca 2700

caggagaggt agcaaaggcc aggcttttct ttggttttct tcaaacatag gtgaaaaaaa 2760

cactgccatt cacaagtcaa ggaacccagg gccagctgga agtgtggagc acacatgctg 2820

tggagcacac atgctgtgga gattgcagtg tgtctgaggt ttgtgtagta gtggaagatt 2880

ttaggtatgt agagcaagtt gaaaatggat tgagactgca tggtggcata aatgagaaat 2940

tgcctgtagc atctagtcta cttgaaggaa gtggagacat aaggagagac aaaaacaggt 3000

ttgtgccata aagtattttt tcaaagacac caagatgtgg taaatgaaaa ttattagttc 3060

acttccctgc tgccatgaaa ctttgcctta agaaggtgct ggattccaag gtttgtaaag 3120

gcatctcggt aaagactgct ttttgaatgc atatgatttt gcatcagcta gactgagttg 3180

attctgacca gacttgatgg ttttaagtcg gaaccgataa attttaaaaa ggagaaaaaa 3240

taatttgacc tagtagtata aaacatgagg ctttaatggt actttgctat gaaaagaaaa 3300

cactgtattc cttatgcaaa acacatgtat ctttcattat ttataagtgg cctctcttag 3360

ctcagttact caattcatac gtagtatttt ttaaaataat tttatatctg tgtaccaccc 3420

catatatttc atattactgt ttcacatgta cagctttcta cttctttgta agaacaccaa 3480

ccaaccaagg tttaagtgat taataggctt gagcaccggg tggcagatgt tctatgcagt 3540

gtggttcaag tttctttgac cgcacttata tgcattgcta atatggaatt taagatacca 3600

tacacagtct ctcatggacc tatctctatt gtagaattat gactttcgtt gtcgaatgac 3660

cactgctgga tgtacctttt tttctgagct ctggtttgcc tttcttgact gtggccatca 3720

ccatgtcacc cacaccagca gcgggaagtc tgttcagccg tcccttgatc cccttcacgg 3780

agatgatata caggtttttg gctcctgtgt tgtcagcaca attgattaca gctcctaccg 3840

gaagacccaa ggaaatccgg aatttcgcac cagaggaccc accacgtcct cgcttcgaca 3900

tcttgaacgc cggaaaaaag aaaaaaggta catccagcag tggtcattcg acaacgaaag 3960

tcataccgta gaaaagatgg cgtgtttctt tattttgaag ataatgcagg agtcatagtg 4020

aacaataaag gcgagatgaa aggttctgcc attacaggac cagtagcaaa ggagtgtgac 4080

gacttgtggc cccggattgc atccaatgct ggcagcattg catgc 4125

<210> 20
<211> 4125
<212> DNA
<213> Homo sapiens

<400> 20
gcatgcaatg ctgccagcat tggatgcaat ccggggccac aagtcgtcac actcctttgc 60

tactggtcct gtaatggcag aacctttcat ctcgccttta ttgttcacta tgactcctgc 120

attatcttca aaataaagaa acacgccatc ttttctacgg tatgactttc gttgtcgaat 180

gaccactgct ggatgtacct ttttttcttt ttccggcgtt caagatgtcg aagcgaggac 240

gtggtgggtc ctctggtgcg aaattccgga tttccttggg tcttccggta ggagctgtaa 300

tcaattgtgc tgacaacaca ggagccaaaa acctgtatat catctccgtg aaggggatca 360

agggacggct gaacagactt cccgctgctg gtgtgggtga catggtgatg gccacagtca 420

agaaaggcaa accagagctc agaaaaaaag gtacatccag cagtggtcat tcgacaacga 480

aagtcataat tctacaatag agataggtcc atgagagact gtgtatggta tcttaaattc 540

catattagca atgcatataa gtgcggtcaa agaaacttga accacactgc atagaacatc 600

tgccacccgg tgctcaagcc tattaatcac ttaaaccttg gttggttggt gttcttacaa 660

agaagtagaa agctgtacat gtgaaacagt aatatgaaat atatggggtg gtacacagat 720

ataaaattat tttaaaaaat actacgtatg aattgagtaa ctgagctaag agaggccact 780

tataaataat gaaagataca tgtgttttgc ataaggaata cagtgttttc ttttcatagc 840

aaagtaccat taaagcctca tgttttatac tactaggtca aattattttt tctccttttt 900

aaaatttatc ggttccgact taaaaccatc aagtctggtc agaatcaact cagtctagct 960
gatgcaaaat catatgcatt caaaaagcag tctttaccga gatgccttta caaaccttgg 1020
aatccagcac cttcttaagg caaagtttca tggcagcagg gaagtgaact aataattttc 1080
atttaccaca tcttggtgtc tttgaaaaaa tactttatgg cacaaacctg tttttgtctc 1140
tccttatgtc tccacttcct tcaagtagac tagatgctac aggcaatttc tcatttatgc 1200
caccatgcag tctcaatcca ttttcaactt gctctacata cctaaaatct tccactacta 1260
cacaaacctc agacacactg caatctccac agcatgtgtg ctccacagca tgtgtgctcc 1320
acacttccag ctggccctgg gttccttgac ttgtgaatgg cagtgttttt ttcacctatg 1380
tttgaagaaa accaaagaaa agcctggcct ttgctacctc tcctgtgcct tctactttta 1440
tctaaatgaa cacactataa tcatgctggt tgcattttcc ttcattctcc tccctactcc 1500
ttccaaaccc ctgattcctc caaatactat atatgccttg gtcatttggg tggtttggat 1560
gctttctgtg gaatgaactg taatatattt ctgctttatt cagagcactt tatattattt 1620
atatttaaat tagcctttta atatggtttg taaataagtt tcctaccaag gtagcatgaa 1680
aacactacac ataaagaaaa aaaatgttaa gaactaaggt attaaccttg ttagaaaact 1740
cttttgatat taacatttaa ggcaagtagt tccttcctta tgatttaata taaaaacatt 1800
tgcaaagaca aggcatttac ttggatttct ttgtcatctc ttcttctttc atccttacat 1860
attgcacttg gttcactgta tctgtctgaa tacactgtgc ttcagaatca ttgtagtact 1920
gtttgtggta caactccctc attaaacatt tcacactatc aggctacaca gcacctatat 1980
cttgtgttta aaggcacttg tagataatac aaattgtagt tacctgtgaa caattgattt 2040
cctttagtat tttggcattt tgatcacttt aatcacagca attccaaagt tttaaattgg 2100
acagtaaatg tgtattttct gtgcaatgtg tatatttgta ttagcattat ctacttaaac 2160
taagtttggt atatcaagta agaaaaatat tattttggtg tttttttta atttcttata 2220
gtcaaaggta tgtttctgcc ttttacataa tgtgacaaag gaatatgttg gtcaaggcaa 2280
tggctgtttc agtgtttcag ctttaacaag aatgctggat tacaggtcct cactttctac 2340
caaggcagta ttcagtgtca ggtgagatgg gttggcctca ggttggaacg ctgctttgat 2400
gtctagtccc tggtccgaaa gtgctgcata gcgactggct gagggccgta ctttttttgg 2460
cttggtgctc tgtgactgct gatgccactg ggtgtaaacg tccagcctgg cagtacagga 2520
cacaatccct gcttcattct tggctgacac agtataccac ccagcatctt cttttgtggc 2580
tccctgaatg agcaggcaga tgtagccgtg gttgtcctgg tgcatgctca ctcggtcagt 2640
gctgtgagtg agtgattcat tttctttctt ccaaaatatc tgaggtggtg gcactcccaa 2700
tacacgacat tccagccgca ctgggtaccc atcagcaact cctgtgtttt ggagcttctc 2760

aataaacaca gggggtttgt gtgcttcttt agcagcaacc acaagctcca ggctgaatga 2820

gttctgtcct gctcggttgg tagctataca tgtgtagatg ccggcatcac gtgacgtgac 2880

tggctctatg atcagagagt gcaccccgtt ctcacgcacc agcatcttgt gagcactgtc 2940

agggcgtacg ggctttccat ctagttgcca gcttagatct ggggttggta acccactgac 3000

tttgcagtcc attctgcaga gttttccttc ttgaacagtc agatctccag gagcctgcaa 3060

gaagtgaggt ctgaagaatc gctcctgaat tggttcattt tcgtctccac tgtcccttga 3120

tctagaacga ggccttctga catgaggatg gcctgaggga gaccggggac ttcgacctct 3180

ttggttgaca gcctgtacca ttagccgtcc agtacaactg atgcggccct gagggtttgc 3240

agccataatt gtataattcc catcatcatc tagggtggag gctgtggtat ggagggagca 3300

ggtcccatcg agatctcttt gaatggtgta gtgatcactc tttggagaga tctgcttccc 3360

atctttaaac caatagatct ttggctttgg atttccagcc actctacatg tgaaagttac 3420

tggcattccc tcaaagatct tgtaatgttt cagcttcatc tcaaagaatg gtgccattcc 3480

gttttccaca ggcacatctc catacgtgaa ttcatcacct gattcatcca caggagacct 3540

ttctagcctg tactctattt cactgatgag tctctgttca cagctggaga ctttgtattc 3600

ctgaatcaag gaaagtttaa gttagcagac tcaatccctt ctgacaaagt gtcacgtctc 3660

tccctgcccc ccaataaggc cctagaagac agtgatattt ttaccataca gcttaaaaac 3720

acagtctgag cccaattcaa ttgcatatga aacaaaaaac atctggaagt gagtttatgt 3780

ggtatttcag attgtcttca agtcctatca tccaaaaaga ggggggtagg aaagatgaat 3840

tttttttcc ttcttttctc tgaggcgagg ggaaacaagg ggaaggaaag aaaaaggcaa 3900

aagcaaagct atgtattcct aaactgttga taacaataaa atataatgga acgctgattt 3960

ccttctgtag ggtaaaagct gatgcagggt ggtacatctc tccacatagt acatgtggca 4020

agcagctgcc atattatata caattatagt aaaagaataa tagaaaaaat ggtactagaa 4080

aaaatggtaa tagaaaaaat agtaaaagaa taatagaaaa aaaaa 4125

<210> 21
<211> 385
<212> PRT
<213> Homo sapiens

<400> 21

Met Ala Pro Phe Phe Glu Met Lys Leu Lys His Tyr Lys Ile Phe Glu
1 5 10 15

Gly Met Pro Val Thr Phe Thr Cys Arg Val Ala Gly Asn Pro Lys Pro
20 25 30

Lys Ile Tyr Trp Phe Lys Asp Gly Lys Gln Ile Ser Pro Lys Ser Asp

35 40 45

His Tyr Thr Ile Gln Arg Asp Leu Asp Gly Thr Cys Ser Leu His Thr
50 55 60

Thr Ala Ser Thr Leu Asp Asp Asp Gly Asn Tyr Thr Ile Met Ala Ala
65 70 75 80

Asn Pro Gln Gly Arg Ile Ser Cys Thr Gly Arg Leu Met Val Gln Ala
85 90 95

Val Asn Gln Arg Gly Arg Ser Pro Arg Ser Pro Ser Gly His Pro His
100 105 110

Val Arg Arg Pro Arg Ser Arg Ser Arg Asp Ser Gly Asp Glu Asn Glu
115 120 125

Pro Ile Gln Glu Arg Phe Phe Arg Pro His Phe Leu Gln Ala Pro Gly
130 135 140

Asp Leu Thr Val Gln Glu Gly Lys Leu Cys Arg Met Asp Cys Lys Val
145 150 155 160

Ser Gly Leu Pro Thr Pro Asp Leu Ser Trp Gln Leu Asp Gly Lys Pro
165 170 175

Val Arg Pro Asp Ser Ala His Lys Met Leu Val Arg Glu Asn Gly Val
180 185 190

His Ser Leu Ile Ile Glu Pro Val Thr Ser Arg Asp Ala Gly Ile Tyr
195 200 205

Thr Cys Ile Ala Thr Asn Arg Ala Gly Gln Asn Ser Phe Ser Leu Glu
210 215 220

Leu Val Val Ala Ala Lys Glu Ala His Lys Pro Pro Val Phe Ile Glu
225 230 235 240

Lys Leu Gln Asn Thr Gly Val Ala Asp Gly Tyr Pro Val Arg Leu Glu
245 250 255

Cys Arg Val Leu Gly Val Pro Pro Pro Gln Ile Phe Trp Lys Lys Glu
260 265 270

Asn Glu Ser Leu Thr His Ser Thr Asp Arg Val Ser Met His Gln Asp
275 280 285

Asn His Gly Tyr Ile Cys Leu Leu Ile Gln Gly Ala Thr Lys Glu Asp
290 295 300

Ala Gly Trp Tyr Thr Val Ser Ala Lys Asn Glu Ala Gly Ile Val Ser
305 310 315 320

Cys Thr Ala Arg Leu Asp Val Tyr Thr Gln Trp His Gln Gln Ser Gln
325 330 335

Ser Thr Lys Pro Lys Lys Val Arg Pro Ser Ala Ser Arg Tyr Ala Ala
340 345 350

Leu Ser Asp Gln Gly Leu Asp Ile Lys Ala Ala Phe Gln Pro Glu Ala
355 360 365

Asn Pro Ser His Leu Thr Leu Asn Thr Ala Leu Val Glu Ser Glu Asp
370 375 380

Leu
385

<210> 22
<211> 191
<212> DNA
<213> Homo sapiens

<400> 22
ttgctttggc ttctgataag attccagcca ctttgtgttt atttattcat tcaacaatta 60

ttgattgtgc acctgctgtg ctcaaggcac ccctcctaag tgctaggaca tgtagaaaac 120

aaagccgtcc cttcgatcac aaaacttact ttctattagg aacaaaaaaa aaaaaaaaaa 180

aaaaaaaacc c 191

<210> 23
<211> 191
<212> DNA
<213> Homo sapiens

<400> 23
gggtttttt ttttttttt ttttttttgt tcctaataga aagtaagttt tgtgatcgaa 60

gggacggctt tgttttctac atgtcctagc acttaggagg ggtgccttga gcacagcagg 120

tgcacaatca ataattgttg aatgaataaa taaacacaaa gtggctggaa tcttatcaga 180

agccaaagca a 191

<210> 24
<211> 32
<212> PRT
<213> Homo sapiens

<400> 24

Leu Leu Trp Leu Leu Ile Arg Phe Gln Pro Leu Cys Val Tyr Leu Phe
1 5 10 15

Ile Gln Gln Leu Leu Ile Val His Leu Leu Cys Ser Arg His Pro Ser
20 25 30

<210> 25
<211> 6450
<212> DNA
<213> Homo sapiens

<400> 25
gagttgtgcc tggagtgatg tttaagccaa tgtcagggca aggcaacagt ccctggccgt 60

cctccagcac ctttgtaatg catatgagct cgggagacca gtacttaaag ttggaggccc 120

gggagcccag gagctggcgg agggcgttcg tcctgggagc tgcacttgct ccgtcgggtc 180

gccggcttca ccggaccgca ggctcccggg gcagggccgg ggccagagct cgcgtgtcgg 240

cgggacatgc gctgcgtcgc ctctaacctc gggctgtgct ctttttccag gtggcccgcc 300

ggtttctgag ccttctgccc tgcggggaca cggtctgcac cctgcccgcg gccacggacc 360

atgaccatga ccctccacac caaagcatct gggatggccc tactgcatca gatccaaggg 420

aacgagctgg agcccctgaa ccgtccgcag ctcaagatcc ccctggagcg gcccctgggc 480

gaggtgtacc tggacagcag caagcccgcc gtgtacaact accccgaggg cgccgcctac 540

gagttcaacg ccgcggccgc cgccaacgcg caggtctacg gtcagaccgg cctcccctac 600

ggccccgggt ctgaggctgc ggcgttcggc tccaacggcc tgggggttt ccccccactc 660

aacagcgtgt ctccgagccc gctgatgcta ctgcacccgc cgccgcagct gtcgcctttc 720

ctgcagcccc acggccagca ggtgccctac tacctggaga acgagcccag cggctacacg 780

gtgcgcgagg ccggcccgcc ggcattctac aggccaaatt cagataatcg acgccagggt 840

ggcagagaaa gattggccag taccaatgac aagggaagta tggctatgga atctgccaag 900

gagactcgct actgtgcagt gtgcaatgac tatgcttcag gctaccatta tggagtctgg 960

tcctgtgagg gctgcaaggc cttcttcaag agaagtattc aaggacataa cgactatatg 1020

tgtccagcca ccaaccagtg caccattgat aaaaacagga ggaagagctg ccaggcctgc 1080

cggctccgca aatgctacga agtgggaatg atgaaaggtg ggatacgaaa agaccgaaga 1140

ggagggagaa tgttgaaaca caagcgccag agagatgatg gggagggcag gggtgaagtg 1200

gggtctgctg gagacatgag agctgccaac ctttggccaa gcccgctcat gatcaaacgc 1260

tctaagaaga acagcctggc cttgtccctg acggccgacc agatggtcag tgccttgttg 1320

gatgctgagc cccccatact ctattccgag tatgatccta ccagaccctt cagtgaagct 1380

tcgatgatgg gcttactgac caacctggca gacagggagc tggttcacat gatcaactgg 1440

gcgaagaggg tgccaggctt tgtggatttg accctccatg atcaggtcca ccttctagaa 1500
tgtgcctggc tagagatcct gatgattggt ctcgtctggc gctccatgga gcacccagtg 1560
aagctactgt ttgctcctaa cttgctcttg gacaggaacc agggaaaatg tgtagagggc 1620
atggtggaga tcttcgacat gctgctggct acatcatctc ggttccgcat gatgaatctg 1680
cagggagagg agtttgtgtg cctcaaatct attattttgc ttaattctgg agtgtacaca 1740
tttctgtcca gcaccctgaa gtctctggaa gagaaggacc atatccaccg agtcctggac 1800
aagatcacag acactttgat ccacctgatg gccaaggcag gcctgaccct gcagcagcag 1860
caccagcggc tggcccagct cctcctcatc ctctcccaca tcaggcacat gagtaacaaa 1920
ggcatggagc atctgtacag catgaagtgc aagaacgtgg tgcccctcta tgacctgctg 1980
ctggagatgc tggacgccca ccgcctacat gcgcccacta gccgtggagg ggcatccgtg 2040
gaggagacgg accaaagcca cttggccact gcgggctcta cttcatcgca ttccttgcaa 2100
aagtattaca tcacggggga ggcagagggt ttccctgcca cagtctgaga gctccctggc 2160
tcccacacgg ttcagataat ccctgctgca ttttaccctc atcatgcacc actttagcca 2220
aattctgtct cctgcataca ctccggcatg catccaacac caatggcttt ctagatgagt 2280
ggccattcat ttgcttgctc agttcttagt ggcacatctt ctgtcttctg ttgggaacag 2340
ccaaagggat tccaaggcta aatctttgta acagctctct ttccccttg ctatgttact 2400
aagcgtgagg attcccgtag ctcttcacag ctgaactcag tctatgggtt ggggctcaga 2460
taactctgtg catttaagct acttgtagag acccaggcct ggagagtaga cattttgcct 2520
ctgataagca ctttttaaat ggctctaaga ataagccaca gcaaagaatt taaagtggct 2580
cctttaattg gtgacttgga gaaagctagg tcaagggttt attatagcac cctcttgtat 2640
tcctatggca atgcatcctt ttatgaaagt ggtacacctt aaagctttta tatgactgta 2700
gcagagtatc tggtgattgt caattcactt ccccctatag gaatacaagg ggccacacag 2760
ggaaggcaga tcccctagtt ggccaagact tattttaact tgatacactg cagattcaga 2820
gtgtcctgaa gctctgcctc tggctttccg gtcatgggtt ccagttaatt catgcctccc 2880
atggacctat ggagagcaac aagttgatct tagttaagtc tccctatatg agggataagt 2940
tcctgatttt tgtttttatt tttgtgttac aaaagaaagc cctccctccc tgaacttgca 3000
gtaaggtcag cttcaggacc tgttccagtg ggcactgtac ttggatcttc ccggcgtgtg 3060
tgtgccttac acaggggtga actgttcact gtggtgatgc atgatgaggg taaatggtag 3120
ttgaaaggag caggggccct ggtgttgcat ttagccctgg ggcatggagc tgaacagtac 3180
ttgtgcagga ttgttgtggc tactagagaa caagagggaa agtagggcag aaactggata 3240
cagttctgag cacagccaga cttgctcagg tggccctgca caggctgcag ctacctagga 3300

acattccttg cagaccccgc attgcctttg ggggtgccct gggatccctg gggtagtcca 3360
gctcttattc atttcccagc gtggccctgg ttggaagaag cagctgtcaa gttgtagaca 3420
gctgtgttcc tacaattggc ccagcaccct ggggcacggg agaagggtgg ggaccgttgc 3480
tgtcactact caggctgact ggggcctggt cagattacgt atgcccttgg tggtttagag 3540
ataatccaaa atcagggttt ggtttgggga agaaaatcct ccccttcct cccccgcccc 3600
gttccctacc gcctccactc ctgccagctc atttccttca atttcctttg acctataggc 3660
taaaaaagaa aggctcattc cagccacagg gcagccttcc ctgggccttt gcttctctag 3720
cacaattatg ggttacttcc tttttcttaa caaaaaagaa tgtttgattt cctctgggtg 3780
accttattgt ctgtaattga aaccctattg agaggtgatg tctgtgttag ccaatgaccc 3840
aggtagctgc tcgggcttct cttggtatgt cttgtttgga aaagtggatt tcattcattt 3900
ctgattgtcc agttaagtga tcaccaaagg actgagaatc tgggagggca aaaaaaaaaa 3960
aaaaagtttt tatgtgcact taaatttggg gacaatttta tgtatctgtg ttaaggatat 4020
gcttaagaac ataattcttt tgttgctgtt tgtttaagaa gcaccttagt ttgtttaaga 4080
agcaccttat atagtataat atatattttt ttgaaattac attgcttgtt tatcagacaa 4140
ttgaatgtag taattctgtt ctggatttaa tttgactggg ttaacatgca aaaaccaagg 4200
aaaaatattt agtttttttt ttttttttg tatacttttc aagctacctt gtcatgtata 4260
cagtcattta tgcctaaagc ctggtgatta ttcatttaaa tgaagatcac atttcatatc 4320
aacttttgta tccacagtag acaaaatagc actaatccag atgcctattg ttggatattg 4380
aatgacagac aatcttatgt agcaaagatt atgcctgaaa aggaaaatta ttcagggcag 4440
ctaattttgc ttttaccaaa atatcagtag taatattttt ggacagtagc taatgggtca 4500
gtgggttctt tttaatgttt atacttagat tttcttttaa aaaaattaaa ataaaacaaa 4560
aaaaatttct aggactagac gatgtaatac cagctaaagc caaacaatta tacagtggaa 4620
ggttttacat tattcatcca atgtgtttct attcatgtta agatactact acatttgaag 4680
tgggcagaga acatcagatg attgaaatgt tcgcccaggg gtctccagca actttggaaa 4740
tctctttgta tttttacttg aagtgccact aatggacagc agatattttc tggctgatgt 4800
tggtattggg tgtaggaaca tgatttaaaa aaaaaactct tgcctctgct ttcccccact 4860
ctgaggcaag ttaaaatgta aaagatgtga tttatctggg gggctcaggt atggtgggga 4920
agtggattca ggaatctggg gaatggcaaa tatattaaga agagtattga aagtatttgg 4980
aggaaaatgg ttaattctgg gtgtgcacca aggttcagta gagtccactt ctgccctgga 5040
gaccacaaat caactagctc catttacagc catttctaaa atggcagctt cagttctaga 5100
gaagaaagaa caacatcagc agtaaagtcc atggaatagc tagtggtctg tgtttctttt 5160
cgccattgcc tagcttgccg taatgattct ataatgccat catgcagcaa ttatgagagg 5220

ctaggtcatc caaagagaag accctatcaa tgtaggttgc aaaatctaac ccctaaggaa 5280

gtgcagtctt tgatttgatt tccctagtaa ccttgcagat atgtttaacc aagccatagc 5340

ccatgccttt tgagggctga acaaataagg gacttactga taatttactt ttgatcacat 5400

taaggtgttc tcaccttgaa atcttataca ctgaaatggc cattgattta ggccactggc 5460

ttagagtact ccttcccctg catgacactg attacaaata ctttcctatt catactttcc 5520

aattatgaga tggactgtgg gtactgggag tgatcactaa caccatagta atgtctaata 5580

ttcacaggca gatctgcttg gggaagctag ttatgtgaaa ggcaaataaa gtcatacagt 5640

agctcaaaag gcaaccataa ttctctttgg tgcaagtctt gggagcgtga tctagattac 5700

actgcaccat tcccaagtta atcccctgaa aacttactct caactggagc aaatgaactt 5760

tggtcccaaa tatccatctt ttcagtagcg ttaattatgc tctgtttcca actgcatttc 5820

ctttccaatt gaattaaagt gtggcctcgt ttttagtcat ttaaaattgt tttctaagta 5880

attgctgcct ctattatggc acttcaattt tgcactgtct tttgagattc aagaaaaatt 5940

tctattcatt tttttgcatc caattgtgcc tgaactttta aaatatgtaa atgctgccat 6000

gttccaaacc catcgtcagt gtgtgtgttt agagctgtgc accctagaaa caacatactt 6060

gtcccatgag caggtgcctg agacacagac ccctttgcat tcacagagag gtcattggtt 6120

atagagactt gaattaataa gtgacattat gccagtttct gttctctcac aggtgataaa 6180

caatgctttt tgtgcactac atactcttca gtgtagagct cttgttttat gggaaaaggc 6240

tcaaatgcca aattgtgttt gatggattaa tatgcccttt tgccgatgca tactattact 6300

gatgtgactc ggttttgtcg cagctttgct ttgtttaatg aaacacactt gtaaacctct 6360

tttgcacttt gaaaaagaat ccagcgggat gctcgagcac ctgtaaacaa ttttctcaac 6420

ctatttgatg ttcaaataaa gaattaaact 6450

<210> 26
<211> 6450
<212> DNA
<213> Homo sapiens

<400> 26
agtttaattc tttatttgaa catcaaatag gttgagaaaa ttgtttacag gtgctcgagc 60

atcccgctgg attctttttc aaagtgcaaa agaggtttac aagtgtgttt cattaaacaa 120

agcaaagctg cgacaaaacc gagtcacatc agtaatagta tgcatcggca aaagggcata 180

ttaatccatc aaacacaatt tggcatttga gccttttccc ataaaacaag agctctacac 240

tgaagagtat gtagtgcaca aaaagcattg tttatcacct gtgagagaac agaaactggc 300

ataatgtcac ttattaattc aagtctctat aaccaatgac ctctctgtga atgcaaaggg 360

gtctgtgtct caggcacctg ctcatgggac aagtatgttg tttctagggt gcacagctct 420

aaacacacac actgacgatg ggtttggaac atggcagcat ttacatattt taaaagttca 480
ggcacaattg gatgcaaaaa aatgaataga aatttttctt gaatctcaaa agacagtgca 540
aaattgaagt gccataatag aggcagcaat tacttagaaa acaattttaa atgactaaaa 600
acgaggccac actttaattc aattggaaag gaaatgcagt tggaaacaga gcataattaa 660
cgctactgaa aagatggata tttgggacca aagttcattt gctccagttg agagtaagtt 720
ttcaggggat taacttggga atggtgcagt gtaatctaga tcacgctccc aagacttgca 780
ccaaagagaa ttatggttgc cttttgagct actgtatgac tttatttgcc tttcacataa 840
ctagcttccc caagcagatc tgcctgtgaa tattagacat tactatggtg ttagtgatca 900
ctcccagtac ccacagtcca tctcataatt ggaaagtatg aataggaaag tatttgtaat 960
cagtgtcatg caggggaagg agtactctaa gccagtggcc taaatcaatg gccatttcag 1020
tgtataagat ttcaaggtga gaacacctta atgtgatcaa aagtaaatta tcagtaagtc 1080
ccttatttgt tcagccctca aaaggcatgg gctatggctt ggttaaacat atctgcaagg 1140
ttactaggga aatcaaatca aagactgcac ttccttaggg gttagatttt gcaacctaca 1200
ttgatagggt cttctctttg gatgacctag cctctcataa ttgctgcatg atggcattat 1260
agaatcatta cggcaagcta ggcaatggcg aaaagaaaca cagaccacta gctattccat 1320
ggactttact gctgatgttg ttctttcttc tctagaactg aagctgccat tttagaaatg 1380
gctgtaaatg gagctagttg atttgtggtc tccagggcag aagtggactc tactgaacct 1440
tggtgcacac ccagaattaa ccattttcct ccaaatactt tcaatactct tcttaatata 1500
tttgccattc cccagattcc tgaatccact tccccaccat acctgagccc cccagataaa 1560
tcacatcttt tacattttaa cttgcctcag agtgggggaa agcagaggca agagtttttt 1620
ttttaaatca tgttcctaca cccaatacca acatcagcca gaaaatatct gctgtccatt 1680
agtggcactt caagtaaaaa tacaaagaga tttccaaagt tgctggagac ccctgggcga 1740
acatttcaat catctgatgt tctctgccca cttcaaatgt agtagtatct taacatgaat 1800
agaaacacat tggatgaata atgtaaaacc ttccactgta taattgtttg gctttagctg 1860
gtattacatc gtctagtcct agaaattttt tttgttttat tttaattttt ttaaaagaaa 1920
atctaagtat aaacattaaa aagaacccac tgacccatta gctactgtcc aaaaatatta 1980
ctactgatat tttggtaaaa gcaaaattag ctgccctgaa taattttcct tttcaggcat 2040
aatctttgct acataagatt gtctgtcatt caatatccaa caataggcat ctggattagt 2100
gctattttgt ctactgtgga tacaaaagtt gatatgaaat gtgatcttca tttaaatgaa 2160
taatcaccag gctttaggca taaatgactg tatacatgac aaggtagctt gaaaagtata 2220
caaaaaaaaa aaaaaaaact aaatattttt ccttggtttt tgcatgttaa cccagtcaaa 2280

ttaaatccag aacagaatta ctacattcaa ttgtctgata aacaagcaat gtaatttcaa 2340

aaaaatatat attatactat ataaggtgct tcttaaacaa actaaggtgc ttcttaaaca 2400

aacagcaaca aaagaattat gttcttaagc atatccttaa cacagataca taaaattgtc 2460

cccaaattta agtgcacata aaaacttttt ttttttttt tgccctccca gattctcagt 2520

cctttggtga tcacttaact ggacaatcag aaatgaatga aatccacttt tccaaacaag 2580

acataccaag agaagcccga gcagctacct gggtcattgg ctaacacaga catcacctct 2640

caatagggtt tcaattacag acaataaggt cacccagagg aaatcaaaca ttctttttg 2700

ttaagaaaaa ggaagtaacc cataattgtg ctagagaagc aaaggcccag ggaaggctgc 2760

cctgtggctg gaatgagcct ttctttttta gcctataggt caaaggaaat tgaaggaaat 2820

gagctggcag gagtggaggc ggtagggaac ggggcggggg aggaaggggg aggattttct 2880

tccccaaacc aaaccctgat tttggattat ctctaaacca ccaagggcat acgtaatctg 2940

accaggcccc agtcagcctg agtagtgaca gcaacggtcc ccacccttct cccgtgcccc 3000

agggtgctgg gccaattgta ggaacacagc tgtctacaac ttgacagctg cttcttccaa 3060

ccagggccac gctgggaaat gaataagagc tggactaccc cagggatccc agggcacccc 3120

caaaggcaat gcggggtctg caaggaatgt tcctaggtag ctgcagcctg tgcagggcca 3180

cctgagcaag tctggctgtg ctcagaactg tatccagttt ctgccctact ttccctcttg 3240

ttctctagta gccacaacaa tcctgcacaa gtactgttca gctccatgcc ccagggctaa 3300

atgcaacacc agggcccctg ctcctttcaa ctaccattta ccctcatcat gcatcaccac 3360

agtgaacagt tcacccctgt gtaaggcaca cacacgccgg gaagatccaa gtacagtgcc 3420

cactggaaca ggtcctgaag ctgaccttac tgcaagttca gggagggagg gctttctttt 3480

gtaacacaaa aataaaaaca aaaatcagga acttatccct catataggga gacttaacta 3540

agatcaactt gttgctctcc ataggtccat gggaggcatg aattaactgg aacccatgac 3600

cggaaagcca gaggcagagc ttcaggacac tctgaatctg cagtgtatca agttaaaata 3660

agtcttggcc aactagggga tctgccttcc ctgtgtggcc ccttgtattc ctatagggggg 3720

aagtgaattg acaatcacca gatactctgc tacagtcata taaaagcttt aaggtgtacc 3780

actttcataa aaggatgcat tgccatagga atacaagagg gtgctataat aaacccttga 3840

cctagctttc tccaagtcac caattaaagg agccacttta aattctttgc tgtggcttat 3900

tcttagagcc atttaaaaag tgcttatcag aggcaaaatg tctactctcc aggcctgggt 3960

ctctacaagt agcttaaatg cacagagtta tctgagcccc aacccataga ctgagttcag 4020

ctgtgaagag ctacgggaat cctcacgctt agtaacatag caaggggggaa agagagctgt 4080

tacaaagatt tagccttgga atccctttgg ctgttcccaa cagaagacag aagatgtgcc 4140

actaagaact gagcaagcaa atgaatggcc actcatctag aaagccattg gtgttggatg 4200

catgccggag tgtatgcagg agacagaatt tggctaaagt ggtgcatgat gagggtaaaa 4260

tgcagcaggg attatctgaa ccgtgtggga gccagggagc tctcagactg tggcagggaa 4320

accctctgcc tcccccgtga tgtaatactt ttgcaaggaa tgcgatgaag tagagcccgc 4380

agtggccaag tggctttggt ccgtctcctc cacggatgcc cctccacggc tagtgggcgc 4440

atgtaggcgg tgggcgtcca gcatctccag cagcaggtca tagaggggca ccacgttctt 4500

gcacttcatg ctgtacagat gctccatgcc tttgttactc atgtgcctga tgtgggagag 4560

gatgaggagg agctgggcca gccgctggtg ctgctgctgc agggtcaggc ctgccttggc 4620

catcaggtgg atcaaagtgt ctgtgatctt gtccaggact cggtggatat ggtccttctc 4680

ttccagagac ttcagggtgc tggacagaaa tgtgtacact ccagaattaa gcaaaataat 4740

agatttgagg cacacaaact cctctccctg cagattcatc atgcggaacc gagatgatgt 4800

agccagcagc atgtcgaaga tctccaccat gccctctaca cattttccct ggttcctgtc 4860

caagagcaag ttaggagcaa acagtagctt cactgggtgc tccatggagc gccagacgag 4920

accaatcatc aggatctcta gccaggcaca ttctagaagg tggacctgat catggagggt 4980

caaatccaca aagcctggca ccctcttcgc ccagttgatc atgtgaacca gctccctgtc 5040

tgccaggttg gtcagtaagc ccatcatcga agcttcactg aagggtctgg taggatcata 5100

ctcggaatag agtatggggg gctcagcatc caacaaggca ctgaccatct ggtcggccgt 5160

cagggacaag gccaggctgt tcttcttaga gcgtttgatc atgagcgggc ttggccaaag 5220

gttggcagct ctcatgtctc cagcagaccc cacttcaccc ctgccctccc catcatctct 5280

ctggcgcttg tgtttcaaca ttctccctcc tcttcggtct tttcgtatcc caccttttcat 5340

cattcccact tcgtagcatt tgcggagccg gcaggcctgg cagctcttcc tcctgttttt 5400

atcaatggtg cactggttgg tggctggaca catatagtcg ttatgtcctt gaatacttct 5460

cttgaagaag gccttgcagc cctcacagga ccagactcca taatggtagc ctgaagcata 5520

gtcattgcac actgcacagt agcgagtctc cttggcagat tccatagcca tacttccctt 5580

gtcattggta ctggccaatc tttctctgcc accctggcgt cgattatctg aatttggcct 5640

gtagaatgcc ggcgggccgg cctcgcgcac cgtgtagccg ctgggctcgt tctccaggta 5700

gtagggcacc tgctggccgt ggggctgcag gaaaggcgac agctgcggcg gcgggtgcag 5760

tagcatcagc gggctcggag acacgctgtt gagtgggggg aaacccccca ggccgttgga 5820

gccgaacgcc gcagcctcag acccggggcc gtaggggagg ccggtctgac cgtagacctg 5880

cgcgttggcg gcggccgcgg cgttgaactc gtaggcggcg ccctcggggt agttgtacac 5940

ggcgggcttg ctgctgtcca ggtacacctc gcccaggggc cgctccaggg ggatcttgag 6000

ctgcggacgg ttcaggggct ccagctcgtt cccttggatc tgatgcagta gggccatccc 6060

agatgctttg gtgtggaggg tcatggtcat ggtccgtggc cgcgggcagg gtgcagaccg 6120

tgtccccgca gggcagaagg ctcagaaacc ggcgggccac ctggaaaaag agcacagccc 6180

gaggttagag gcgacgcagc gcatgtcccg ccgacacgcg agctctggcc ccggccctgc 6240

cccgggagcc tgcggtccgg tgaagccggc gacccgacgg agcaagtgca gctcccagga 6300

cgaacgccct ccgccagctc ctgggctccc gggcctccaa ctttaagtac tggtctcccg 6360

agctcatatg cattacaaag gtgctggagg acggccaggg actgttgcct tgccctgaca 6420

ttggcttaaa catcactcca ggcacaactc 6450

<210> 27
<211> 595
<212> PRT
<213> Homo sapiens

<400> 27

Met Thr Met Thr Leu His Thr Lys Ala Ser Gly Met Ala Leu Leu His
1 5 10 15

Gln Ile Gln Gly Asn Glu Leu Glu Pro Leu Asn Arg Pro Gln Leu Lys
20 25 30

Ile Pro Leu Glu Arg Pro Leu Gly Glu Val Tyr Leu Asp Ser Ser Lys
35 40 45

Pro Ala Val Tyr Asn Tyr Pro Glu Gly Ala Ala Tyr Glu Phe Asn Ala
50 55 60

Ala Ala Ala Ala Asn Ala Gln Val Tyr Gly Gln Thr Gly Leu Pro Tyr
65 70 75 80

Gly Pro Gly Ser Glu Ala Ala Ala Phe Gly Ser Asn Gly Leu Gly Gly
85 90 95

Phe Pro Pro Leu Asn Ser Val Ser Pro Ser Pro Leu Met Leu Leu His
100 105 110

Pro Pro Pro Gln Leu Ser Pro Phe Leu Gln Pro His Gly Gln Gln Val
115 120 125

Pro Tyr Tyr Leu Glu Asn Glu Pro Ser Gly Tyr Thr Val Arg Glu Ala
130 135 140

Gly Pro Pro Ala Phe Tyr Arg Pro Asn Ser Asp Asn Arg Arg Gln Gly
145 150 155 160

Gly Arg Glu Arg Leu Ala Ser Thr Asn Asp Lys Gly Ser Met Ala Met

165 170 175

Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala Val Cys Asn Asp Tyr Ala
180 185 190

Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys Lys Ala Phe
195 200 205

Phe Lys Arg Ser Ile Gln Gly His Asn Asp Tyr Met Cys Pro Ala Thr
210 215 220

Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys Ser Cys Gln Ala Cys
225 230 235 240

Arg Leu Arg Lys Cys Tyr Glu Val Gly Met Met Lys Gly Gly Ile Arg
245 250 255

Lys Asp Arg Arg Gly Gly Arg Met Leu Lys His Lys Arg Gln Arg Asp
260 265 270

Asp Gly Glu Gly Arg Gly Glu Val Gly Ser Ala Gly Asp Met Arg Ala
275 280 285

Ala Asn Leu Trp Pro Ser Pro Leu Met Ile Lys Arg Ser Lys Lys Asn
290 295 300

Ser Leu Ala Leu Ser Leu Thr Ala Asp Gln Met Val Ser Ala Leu Leu
305 310 315 320

Asp Ala Glu Pro Pro Ile Leu Tyr Ser Glu Tyr Asp Pro Thr Arg Pro
325 330 335

Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr Asn Leu Ala Asp Arg
340 345 350

Glu Leu Val His Met Ile Asn Trp Ala Lys Arg Val Pro Gly Phe Val
355 360 365

Asp Leu Thr Leu His Asp Gln Val His Leu Leu Glu Cys Ala Trp Leu
370 375 380

Glu Ile Leu Met Ile Gly Leu Val Trp Arg Ser Met Glu His Pro Val
385 390 395 400

Lys Leu Leu Phe Ala Pro Asn Leu Leu Leu Asp Arg Asn Gln Gly Lys
405 410 415

Cys Val Glu Gly Met Val Glu Ile Phe Asp Met Leu Leu Ala Thr Ser
420 425 430

Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu Glu Phe Val Cys Leu
435 440 445

Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr Thr Phe Leu Ser Ser
450 455 460

Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile His Arg Val Leu Asp
465 470 475 480

Lys Ile Thr Asp Thr Leu Ile His Leu Met Ala Lys Ala Gly Leu Thr
485 490 495

Leu Gln Gln Gln His Gln Arg Leu Ala Gln Leu Leu Leu Ile Leu Ser
500 505 510

His Ile Arg His Met Ser Asn Lys Gly Met Glu His Leu Tyr Ser Met
515 520 525

Lys Cys Lys Asn Val Val Pro Leu Tyr Asp Leu Leu Leu Glu Met Leu
530 535 540

Asp Ala His Arg Leu His Ala Pro Thr Ser Arg Gly Gly Ala Ser Val
545 550 555 560

Glu Glu Thr Asp Gln Ser His Leu Ala Thr Ala Gly Ser Thr Ser Ser
565 570 575

His Ser Leu Gln Lys Tyr Tyr Ile Thr Gly Glu Ala Glu Gly Phe Pro
580 585 590

Ala Thr Val
595

<210> 28
<211> 1044
<212> DNA
<213> Homo sapiens

<400> 28
ggaatgatga aaggtgggat acgaaaagac cgaagaggag ggagaatgtt gaaacacaag 60

cgccagagag atgatgggga gggcaggggt gaagtggggt ctgctggaga catgagagct 120

gccaaccttt ggccaagccc gctcatgatc aaacgctcta agaagaacag cctggccttg 180

tccctgacgg ccgaccagat ggtcagtgcc ttgttggatg ctgagccccc catactctat 240

tccgagtatg atcctaccag acccttcagt gaagcttcga tgatgggctt actgaccaac 300

ctggcagaca gggagctggt tcacatgatc aactgggcga agagggtgcc aggctttgtg 360

gatttgaccc tccatgatca ggtccacctt ctagaatgtg cctggctaga gatcctgatg 420

attggtctcg tctggcgctc catggagcac ccagtgaagc tactgtttgc tcctaacttg 480

ctcttggaca ggaaccaggg aaaatgtgta gagggcatgg tggagatctt cgacatgctg 540

ctggctacat catctcggtt ccgcatgatg aatctgcagg gagaggagtt tgtgtgcctc 600

aaatctatta ttttgcttaa ttctggagtg tacacatttc tgtccagcac cctgaagtct 660

ctggaagaga aggaccatat ccaccgagtc ctggacaaga tcacagacac tttgatccac 720

ctgatggcca aggcaggcct gaccctgcag cagcagcacc agcggctggc ccagctcctc 780

ctcatcctct cccacatcag gcacatgagt aacaaaggca tggagcatct gtacagcatg 840

aagtgcaaga acgtggtgcc cctctatgac ctgctgctgg agatgctgga cgcccaccgc 900

ctacatgcgc ccactagccg tggaggggca tccgtggagg agacggacca aagccacttg 960

gccactgcgg gctctacttc atcgcattcc ttgcaaaagt attacatcac gggggaggca 1020

gagggtttcc ctgccacagt ctga 1044

<210> 29
<211> 1044
<212> DNA
<213> Homo sapiens

<400> 29
tcagactgtg gcagggaaac cctctgcctc ccccgtgatg taatactttt gcaaggaatg 60

cgatgaagta gagcccgcag tggccaagtg gctttggtcc gtctcctcca cggatgcccc 120

tccacggcta gtgggcgcat gtaggcggtg ggcgtccagc atctccagca gcaggtcata 180

gaggggcacc acgttcttgc acttcatgct gtacagatgc tccatgcctt tgttactcat 240

gtgcctgatg tgggagagga tgaggaggag ctgggccagc cgctggtgct gctgctgcag 300

ggtcaggcct gccttggcca tcaggtggat caaagtgtct gtgatcttgt ccaggactcg 360

gtggatatgg tccttctctt ccagagactt cagggtgctg gacagaaatg tgtacactcc 420

agaattaagc aaaataatag atttgaggca cacaaactcc tctccctgca gattcatcat 480

gcggaaccga gatgatgtag ccagcagcat gtcgaagatc tccaccatgc cctctacaca 540

ttttccctgg ttcctgtcca agagcaagtt aggagcaaac agtagcttca ctgggtgctc 600

catggagcgc cagacgagac caatcatcag gatctctagc caggcacatt ctagaaggtg 660

gacctgatca tggagggtca aatccacaaa gcctggcacc ctcttcgccc agttgatcat 720

gtgaaccagc tccctgtctg ccaggttggt cagtaagccc atcatcgaag cttcactgaa 780

gggtctggta ggatcatact cggaatagag tatgggggc tcagcatcca acaaggcact 840

gaccatctgg tcggccgtca gggacaaggc caggctgttc ttcttagagc gtttgatcat 900

gagcgggctt ggccaaaggt tggcagctct catgtctcca gcagacccca cttcacccct 960

gccctcccca tcatctctct ggcgcttgtg tttcaacatt ctccctcctc ttcggtcttt 1020

tcgtatccca cctttcatca ttcc 1044

<210> 30
<211> 347
<212> PRT
<213> Homo sapiens

<400> 30

Gly Met Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met
1 5 10 15

Leu Lys His Lys Arg Gln Arg Asp Asp Gly Glu Gly Arg Gly Glu Val
20 25 30

Gly Ser Ala Gly Asp Met Arg Ala Ala Asn Leu Trp Pro Ser Pro Leu
35 40 45

Met Ile Lys Arg Ser Lys Lys Asn Ser Leu Ala Leu Ser Leu Thr Ala
50 55 60

Asp Gln Met Val Ser Ala Leu Leu Asp Ala Glu Pro Pro Ile Leu Tyr
65 70 75 80

Ser Glu Tyr Asp Pro Thr Arg Pro Phe Ser Glu Ala Ser Met Met Gly
85 90 95

Leu Leu Thr Asn Leu Ala Asp Arg Glu Leu Val His Met Ile Asn Trp
100 105 110

Ala Lys Arg Val Pro Gly Phe Val Asp Leu Thr Leu His Asp Gln Val
115 120 125

His Leu Leu Glu Cys Ala Trp Leu Glu Ile Leu Met Ile Gly Leu Val
130 135 140

Trp Arg Ser Met Glu His Pro Val Lys Leu Leu Phe Ala Pro Asn Leu
145 150 155 160

Leu Leu Asp Arg Asn Gln Gly Lys Cys Val Glu Gly Met Val Glu Ile
165 170 175

Phe Asp Met Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu
180 185 190

```
Gln Gly Glu Glu Phe Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser
        195                 200                 205

Gly Val Tyr Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys
    210                 215                 220

Asp His Ile His Arg Val Leu Asp Lys Ile Thr Asp Thr Leu Ile His
225                 230                 235                 240

Leu Met Ala Lys Ala Gly Leu Thr Leu Gln Gln Gln His Gln Arg Leu
                245                 250                 255

Ala Gln Leu Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys
            260                 265                 270

Gly Met Glu His Leu Tyr Ser Met Lys Cys Lys Asn Val Val Pro Leu
        275                 280                 285

Tyr Asp Leu Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Pro
    290                 295                 300

Thr Ser Arg Gly Gly Ala Ser Val Glu Glu Thr Asp Gln Ser His Leu
305                 310                 315                 320

Ala Thr Ala Gly Ser Thr Ser Ser His Ser Leu Gln Lys Tyr Tyr Ile
                325                 330                 335

Thr Gly Glu Ala Glu Gly Phe Pro Ala Thr Val
            340                 345
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2002 (12.09.2002)

PCT

(10) International Publication Number
WO 02/070699 A3

(51) International Patent Classification[7]: C12N 15/12, 5/10, C07K 14/47, 16/18, G01N 33/68, A01K 67/027

(21) International Application Number: PCT/EP02/02189

(22) International Filing Date: 28 February 2002 (28.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01105062.2 1 March 2001 (01.03.2001) EP

(71) Applicant *(for all designated States except US)*: **LION BIOSCIENCE AG** [DE/DE]; Im Neuenheimer Feld 515, 69120 Heidelberg (DE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ALBERS, Michael** [DE/DE]; Obere Schulstrasse 3, 69221 Dossenheim (DE). **ELLWANGER, Silvia** [DE/DE]; Theodor-Heuss-Strasse 4, 69181 Leimen (DE). **LÖSER, Eva** [DE/DE]; Biethstrasse 12, 69121 Heidelberg (DE). **KÖGL, Manfred** [AT/DE]; Hauptstrasse 131/4, 69214 Eppelheim (DE).

(74) Agents: **GODDAR, Heinz** et al.; Boehmert & Boehmert, Pettenkoferstrasse 20-22, 80336 München (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
13 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/070699 A3

(54) Title: COFACTORS OF THE ESTROGEN RECEPTOR ALPHA AND METHODS OF USE

(57) Abstract: The present invention relates to novel cofactors of the estrogen receptor alpha which are designated CF16, CF17, CF18, CF19, CF40, CF41, CF42 and CF43 and in particular to the isolated nucleic acid sequences encoding these cofactors and the isolated polypeptides thereof. The invention further relates to processes for isolating and/or producing the nucleic acids or the proteins as well as methods of use of these cofactors, such as inhibiting or activating the binding of the cofactors to the estrogen receptor alpha.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/02189

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C12N5/10 C07K14/47 C07K16/18 G01N33/68
A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K G01N A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
SEQUENCE SEARCH, BIOSIS, EPO-Internal, EMBASE, WPI Data, PAJ, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL [Online] Database accession no. AK023173 XP002234411 abstract ---	1-12
X	DUTERTRE MARTIN ET AL: "Molecular mechanisms of selective estrogen receptor modulator (SERM) action." JOURNAL OF PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS, vol. 295, no. 2, November 2000 (2000-11), pages 431-437, XP002234407 ISSN: 0022-3565	1-14
Y	page 431 page 434 -page 436 ---	15-25
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 March 2003	03. 06. 2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Schneider, P

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 99 50664 A (GLAXO GROUP LTD ;HART CHARLES PRARAY (US); NORTHROP JEFFREY PAUL () 7 October 1999 (1999-10-07) page 1 -page 12	15-25
X	KATZENELLENBOGEN B S ET AL: "ESTROGEN RECEPTOR TRANSCRIPTION AND TRANSACTIVATION ESTROGEN RECEPTOR ALPHA AND ESTROGEN RECEPTOR BETA: REGULATION BY SELECTIVE ESTROGEN RECEPTOR MODULATORS AND IMPORTANCE IN BREAST CANCER" BREAST CANCER RESEARCH, CURRENT SCIENCE, LONDON, GB, vol. 2, no. 5, 2000, pages 335-344, XP001148108 ISSN: 1465-5411 page 339 -page 340	1-14
X	LANZ R B ET AL: "A STEROID RECEPTOR COACTIVATOR, SRA, FUNCTIONS AS AN RNA AND IS PRESENT IN AN SRC-1 COMPLEX" CELL, CELL PRESS, CAMBRIDGE, NA, US, vol. 97, 2 April 1999 (1999-04-02), pages 16-27, XP002926500 ISSN: 0092-8674 page 17 page 23 -page 26	16-22
X	SHANG YONGFENG ET AL: "Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription." CELL, vol. 103, no. 6, 8 December 2000 (2000-12-08), pages 843-852, XP002234408 ISSN: 0092-8674 page 843 page 846	16-20
A	MCKENNA NEIL J ET AL: "Nuclear receptor coregulators: Cellular and molecular biology." ENDOCRINE REVIEWS, vol. 20, no. 3, June 1999 (1999-06), pages 321-344, XP002234409 ISSN: 0163-769X page 338 -page 339	

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	PAVAO M ET AL: "Estrogen receptor antibodies: specificity and utility in detection, localization and analyses of estrogen receptor alpha and beta" STEROIDS, BUTTERWORTH-HEINEMANN, STONEHAM, MA, US, vol. 66, no. 1, 1 January 2001 (2001-01-01), pages 1-16, XP004220761 ISSN: 0039-128X the whole document ---	
A	KATZENELLENBOGEN BENITA S ET AL: "Molecular mechanisms of estrogen action: Selective ligands and receptor pharmacology." JOURNAL OF STEROID BIOCHEMISTRY AND MOLECULAR BIOLOGY, vol. 74, no. 5, 2000, pages 279-285, XP002234410 ISSN: 0960-0760 page 282 ---	
A	BARKHEM T ET AL: "DIFFERENTIAL RESPONSE OF ESTROGEN RECEPTOR ALPHA AND ESTROGEN RECEPTOR BETA TO PARTIAL ESTROGEN AGONISTS/ANTAGONISTS" MOLECULAR PHARMACOLOGY, BALTIMORE, MD, US, vol. 54, no. 1, July 1998 (1998-07), pages 105-112, XP000978331 ISSN: 0026-895X cited in the application the whole document -----	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 02/02189

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 18 and 25 are directed to or may comprise a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: 17
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-25 partially

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 17

Present claims 17, 18 and 19 relate to a product/compound defined by reference to a desirable characteristic or property, namely a binding agent and a compound identified by the method of claim 16.

The claims cover all products/compounds having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for only a very limited number of such products/compounds. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the product/compound by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts relating to the products/compounds antibody, antisense and ribozyme, RNA and steroids as mentioned in the description at pages 2, 3, 8 and 32 to 39. As no compound identified by the method of claim 16 has been disclosed, claim 17 was not searched.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-25 partially

   A cofactor of human estrogen receptor alpha as defined by SEQ ID NOs: 1, 2 and 3 (CF16) and related to said sequences a vector, a host cell, a proteinous complex additionally comprising the receptor, a method to screen for binding and modulating compounds of said cofactor or said receptor, a method for modulating the activity of said cofactor and a method of treament.

2. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 4, 5 and 6 (CF17)

3. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 7, 8 and 9 (CF18)

4. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 10, 11 and 12 (CF19)

5. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 13, 14 and 15 (CF40)

6. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 16, 17 and 18 (CF41)

7. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 19, 20 and 21 (CF42)

8. Claims: 1-25 partially

   Same as invention no. 1 but directed to SEQ ID NOs: 22, 23 and 24 (CF43)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/EP 02/02189

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9950664 A	07-10-1999	US 6410245 B1	25-06-2002
		AU 3547999 A	18-10-1999
		CA 2326573 A1	07-10-1999
		EP 1070254 A1	24-01-2001
		JP 2002510051 T	02-04-2002
		WO 9950664 A1	07-10-1999